REVIEW

Cognitive computing and eScience in health and life science research: artificial intelligence and obesity intervention programs

Thomas Marshall[1]* , Tiffiany Champagne-Langabeer[2], Darla Castelli[3] and Deanna Hoelscher[4]

Abstract

Objective: To present research models based on artificial intelligence and discuss the concept of cognitive computing and eScience as disruptive factors in health and life science research methodologies.

Methods: The paper identifies big data as a catalyst to innovation and the development of artificial intelligence, presents a framework for computer-supported human problem solving and describes a transformation of research support models. This framework includes traditional computer support; federated cognition using machine learning and cognitive agents to augment human intelligence; and a semi-autonomous/autonomous cognitive model, based on deep machine learning, which supports eScience.

Results: The paper provides a forward view of the impact of artificial intelligence on our human–computer support and research methods in health and life science research.

Conclusions: By augmenting or amplifying human task performance with artificial intelligence, cognitive computing and eScience research models are discussed as novel and innovative systems for developing more effective adaptive obesity intervention programs.

Keywords: Cognitive computing, Artificial intelligence, Obesity intervention

Introduction

Medical science has, over centuries, established a deliberate and lengthy process of generating theory based on incremental and isolated (episodic) points of discovery. The cycle time in the health and life sciences, which is the period between discovery and practice based on accepted theory, is a critical factor in our society [1]. This cycle time for medical discovery to practice has been decreasing at a significant pace [2, p. 60]. Sciences rely on theory development based on hypothetico-deductive modeling, which involves the process of establishing a priori hypotheses and testing for acceptance or rejection using statistical analysis. As a result of the explosion of data available, many research disciplines previously based on the process of a priori theory development are

undergoing a radical transformation, moving from an a priori model to a data-driven *posterior* model. Jim Grey's Fourth Paradigm [2, foreword, xvii] refers to this phenomenon of computationally-based research as eScience. eScience, per Grey, is a method of data exploration that unifies theory, experiment, and simulation. The Fourth Paradigm shift suggests a new era of research that is based on massive scales of ubiquitous/abundant data that are computationally processed and stored prior to human intervention. eScience as a methodology of scientific discovery in health and life sciences may be tremendously disruptive to the status quo and result in significant benefits to society [2].

Artificial intelligence (AI), foundational in the development of eScience, is the top technology recently identified as disruptive within healthcare, and is expected to "assist healthcare practitioners in using medical knowledge … [and deliver] clinically relevant, real-time, quality information." [3]. By 2025, AI is expected to be "implemented

*Correspondence: thomasemarshall@gmail.com
[1] Auburn University, Auburn, AL, USA
Full list of author information is available at the end of the article

Marshall *et al. Health Inf Sci Syst (2017) 5:13*

Page 2 of 11

in 90% of US and 60% of global hospitals and insurance companies." Motivators for AI implementation include improved patient outcomes, reduced treatment costs, and patient-centric treatment plans. AI supports treatment strategies of directly observed therapy (DOT) and can improve patient outcomes based on automated patient guidance and engagement solutions to improve treatment program adherence [1, 3].

Cognitive computing (CC) is a general perspective presenting AI as computing capabilities that: can be applied to augment/amplify human cognition, scales well, and replicates human expertise in cognitive-task performance. Cognitive computing includes AI technologies such as machine learning, natural language understanding, speech and image recognition, conversant human interface, distributed and high-performance computing [4]. Within healthcare, CC is particularly important because it allows us to give meaning to who, what, where, and when types of questions. CC allows data mining of both key elements—the question and possible solutions—to create alternative, likely even surprising solutions. In this manner, CC strategies can be applied to improve the delivery of personalized health intervention programs, including life-style/behavioral change management, reducing cost and improving program success [5].

Cognitive computing and eScience redefine the paradigm for research and practice of health and life sciences from an a priori dominated methodology to a data-driven *posterior* approach to augment intelligence [6]. This study presents CC and eScience paradigms for implementing data-driven approaches to developing intelligent obesity intervention programs within an adolescent population. Additionally, a transformation path from traditional approaches of computer support for research to an eScience research paradigm based on cognitive computing and deep machine learning is presented.

Task domain—obesity epidemic

Childhood obesity, considered at epidemic levels in the United States, is a chronic health condition impacting both psychological and physical health. The implications of comorbidities associated with obesity are individually and societally significant. Morbidly obese individuals experience diminished life span and quality of life. As a societal responsibility, the costs of health care and the lost productivity associated with obesity are substantial. The purpose of this research is to develop a more effective system for health intervention addressing obesity and related complications within adolescent populations. This study addresses the epidemic problem of obesity among children in a novel and innovative manner by using CC and eScience as research paradigms. The significance of this research is that, with a data-driven

approach including previous program efficacies, the research paradigm uses machine learning to establish a rich knowledge-based system for developing personalized obesity intervention programs to support life-style change. With deep machine learning, this technology implements a research paradigm to support multiphase optimization strategies (MOST) allowing the researchers to determine the components of the program and their optimal dosage to design effective interventions [7–9]. Of paramount importance in behavioral change interventions, MOST treatment programs hold the promised answer to the very critical question of "What Works?" [7].

Technology and methodological transformation

Technological support in the health sciences can be dramatically enhanced by changing from pre-programmed applications to collaborative systems based on deep machine learning, where CC and eScience paradigms implement continuous improvement for more effective obesity intervention design. This study begins with a framework for computer-supported human problem solving, followed by a presentation of a computing transformation from: current computer-based support for human interpretation, to augmented human intelligence from cognitive computing based on machine learning principles, to an eScience research paradigm employing deep machine learning to design adaptive and continuously-improving intervention programs [10, 11].

The efficacy of using technology in a persuasive manner to influence health-related behaviors has been recognized in previous research [11–13]. As with mass-customization models, persuasive technology provides the experience of personalization within given environmental constraints [12]. In the traditional approach, health modification programs are initially designed by human authorities with an objective to be implemented in a manner to improve wellness. This is a common protocol in the health and life sciences: sub-population classification, key factor identification, and intervention program development. In addition, these intervention programs should be capable of scaling to extend their reach to meet the emerging healthcare demands of the population. In the CC and eScience paradigms, the human-developed health modification program, as the initial level of knowledge, becomes the foundation for employing supervised, unsupervised, and reinforced machine learning techniques to create adaptive cognitive agents. With deeper machine learning, the knowledge-based system supports discovering, confirming, and implementing new knowledge in an effort of sustainable self-improvement. Because machine learning is a representation of rapid discovery, the system can succinctly ingest new knowledge and use the accumulated

Marshall *et al. Health Inf Sci Syst (2017) 5:13*

Page 3 of 11

knowledge as a basis for developing more effective and personalized intervention programs. In this manner, CC and eScience impact our healthcare system by reducing the cycle time from discovery to effective intervention program implementation [5]. By implementing MOST research paradigms, CC and eScience can quickly identify program components, their relevancy, and requisite dosage for optimized intervention effect [3, 5–7].

As a parallel in health interventions, the deeper machine learning concept can be applied to move beyond traditional interventions grounded in a priori assumptions, to multicomponent, customizable interventions. Methodologically, rigorous scientific examination of intervention effects has largely focused on the efficacy and fidelity of programming, which is compared to a control or non-intervention condition(s). This approach takes time and may only be able to quantify small portions of the possible intervention effects in a targeted population [14]. Given the potential of systematic rapid learning among cognitive computing and eScience, the interest in multicomponent interventions that are responsive to individual participant needs and can be rigorously examined is gaining momentum [15].

Summarizing, the computer–human problem-solving framework along with the CC and eScience systems will impact delivery and research in the health and life sciences. The CC and eScience systems previously introduced will be discussed in greater detail in the following sections. After this general discussion, the system models are presented, and descriptions of potential applications are discussed. The paper concludes with a summation of the relationship of CC and eScience with more traditional research methodologies and the potential for cognitive computing and eScience paradigms in the health and life sciences to impact health care delivery.

Data as a catalyst for thinking outside of the box

The transformation of research methodologies to a posteriori approaches is based, in part, on the new characteristics of big data commonly referred to as the three Vs (3Vs): variety, velocity, and volume [1]. Recent studies have estimated that the amount of data, structured and unstructured, created in the last two years is equivalent to the cumulative volume of all prior years, with over 80% estimated to be unstructured. Sensors, instrumentation, and the internet of things (IOT) are creating constant streams of data to be ingested into repositories for future use. In healthcare, data is being created at phenomenal rates [1]; as an example, IBM Explorys has clinical data on over 50 million people [5], offering capabilities for rapid learning based on rates of data ingestion and processing previously unattainable [16]. Devices from the internet of things (IOT), including social media-based systems, can capture and deliver valuable data for the

system to ingest and process. However, in obesity-related healthcare, the question remains as to how best to determine relevant data that can be applied in a manner to improve intervention effectiveness.

Data-driven intelligent systems

In a 2009 IEEE article regarding intelligent systems, Google executives reference Eugene Wigner's seminal article, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences, 1960," [17] as a cornerstone for much of their work in the behavioral science domain [18]. In their 2009 article entitled, "The Unreasonable Effectiveness of Data," the authors address intelligent systems and model complexity as it relates to the scale of data observation points, especially within behavioral sciences [18]. One observation made by the authors is that mathematical models are exceptionally effective in the natural sciences with inherently large datasets that are computational in format. On the other hand, the behavioral sciences have lacked such large-scale datasets in computational format. Behavioral datasets are becoming available in various forms in a variety of formats, with varying degrees of computational power. The creation rate of individual behavioral data, availability of that data in society, and the level of under-utilized data is expected to continue its rapid increase [19, 20].

As an example of under-utilized data, school districts produce and maintain large databases containing demographic, contextual, environmental, medical and academic information for each student who has attended at least one day of school. The potential for use and application of such data to enhance the educational experience is grossly underdeveloped. Given the potential of CC and eScience research with the accumulating corpus of a behavioral data lake, new effective intelligent systems are revolutionizing the world and changing life as we know it [18, 20].

Human–computer problem solving

Newell and Newell and Simon, investigating computer support for human problem solving, delineated the components to consist of: data, procedures, goals and constraints, and flexible strategies [21, 22]. Luconi, Malone, and Scott-Morton extended this research with an information systems framework that integrates the problem solving component into system type (see Fig. 1) [11]. The Luconi et al. model details IS type by problem solving component and unit of control (human or computer). More structured problem types (I and II) focus on data, procedures and goals and constraints, with humans controlling flexible strategies. For less structured problem types (III and IV), the Luconi et al. framework includes the application of AI in problem solving individually

Marshall *et al. Health Inf Sci Syst (2017) 5:13*

Page 4 of 11
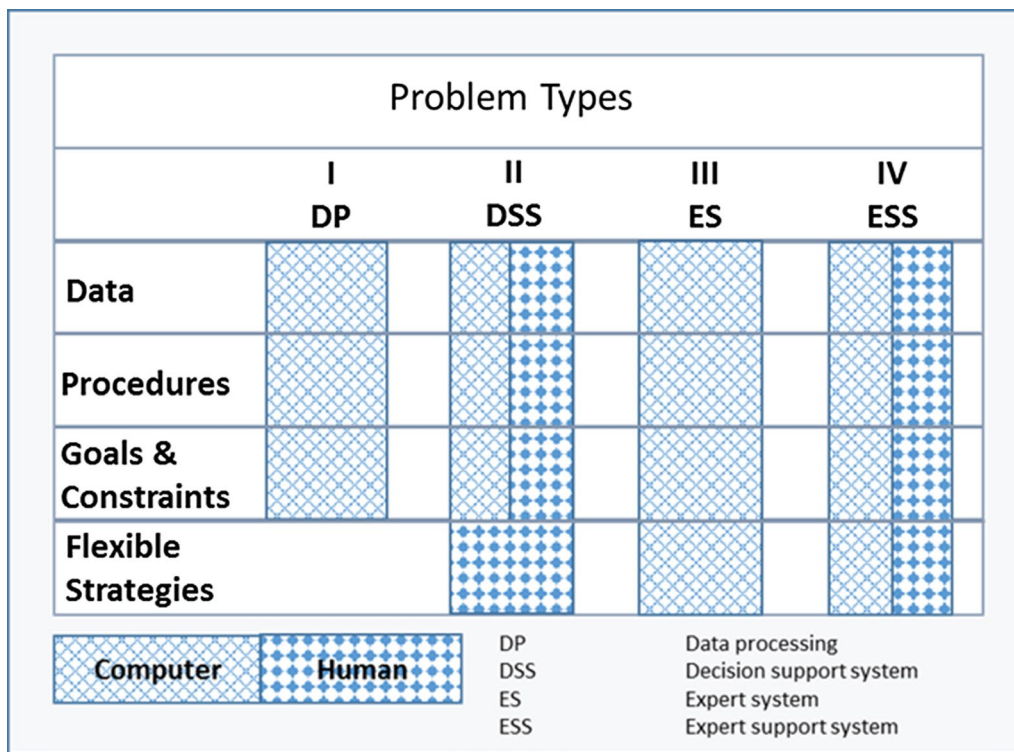
Fig. 1 Luconi, Malone, Scott-Morton IS framework. Adapted from expert systems: The next challenge for managers, Luconi et al. [11]

(Type III) and in collaboration with humans (Type IV). Therefore, the Luconi et al. model views AI, not as an absolute replacement for humans, but as a resource to support human problem solving in a collaborative effort. Based on Newell and Simon's work, the Luconi et al. model is designed to specifically accommodate unstructured decision-making by capitalizing on the power of AI [11, 22]. More structured problems, Types I and II, rely on more traditional technologies generally focusing on data, procedures, and goals and constraints. Given the volume and unstructured nature of data being generated in healthcare, the potential for using AI in problem solving has never been greater. The Luconi et al. framework can be applied to assign tasks into problem types and then relate technology to the problem solving component (data, procedure, goals and constraints, and flexible strategies) to most effectively support human task performance. This model is exceptionally appropriate for the big data era and the application of AI to enhance human performance through cognitive task automation.

Machine learning and other technological advancements are automating tasks that were once engineering bottlenecks, making cognitive task automation both possible and practical [23]. Task automation is no longer limited to routine problem types (I and II) but now includes many cognitive computing applications. One significant impact of artificial intelligence on human problem solving, through cognitive computing and eScience, is to support and automate flexible strategies. Cognitive task automation of flexible strategies holds great promise in complex problem-solving scenarios, such as health and medicine. Cognitive computing and eScience architectures in health and medicine are transformational disruptors to research and methodology if strategically applied. The following sections of the paper detail a transformation of computer support models beginning with traditional computer support for internalized cognition, followed by cognitive computing for federated cognition of best practices, and culminating with semi-autonomous and autonomous cognitive models to support eScience in health and medicine.

Traditional computing

Traditional computer support for human problem solving generally presents the computer as a tool with a primary focus on data, procedures and, perhaps, goals and constraints. In this manner, users internally process information they receive from the computer through interactive sessions, see Fig. 2. Given problem-types I and II, problem type II has the human controlling the activities identified in flexible strategies and all other problem-solving components are shared between the human and
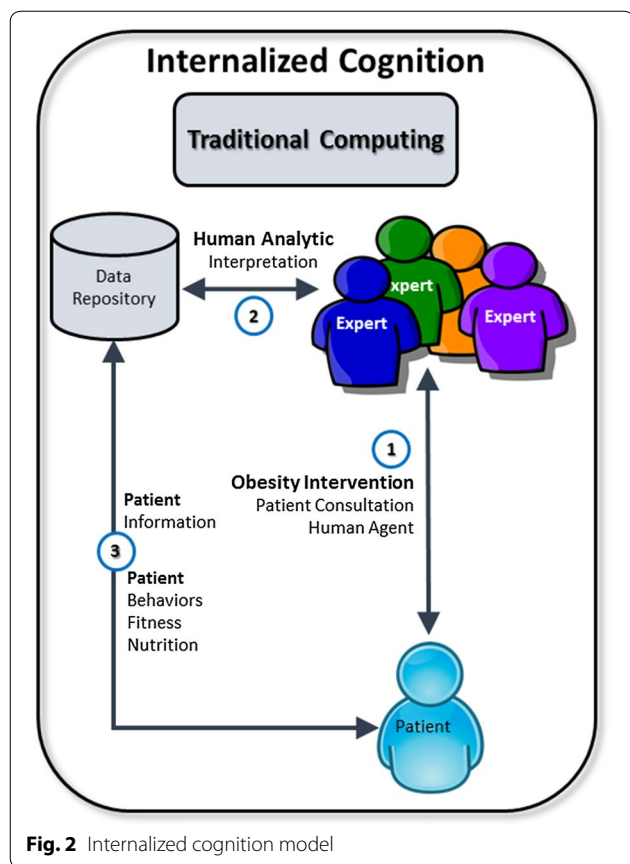
Marshall *et al. Health Inf Sci Syst* (2017) 5:13

Page 5 of 11

Fig. 2 Internalized cognition model

computer. Problem solving in this model relies on and is limited in certain instances by human internal cognitive processing for task execution. Based on AI, CC and eScience models offering augmented/amplified human intelligence are extending the capabilities of collaborative problem solving.

Cognitive computing

Cognitive computing and eScience, based on AI, allow the computer to serve as a source of cognition in problem solving. Cognitive computing encapsulates knowledge in the system allowing for federating, democratizing, and automating cognitive processes. Federating and democratizing cognitive processes, perhaps as best cognitive practices, is facilitated through interfaces built on intelligent agents termed cognitive agents. Cognitive agents rely on AI capabilities in natural language understanding and conversational services to deliver natural communications with humans (Fig. 3, Refs. [5, 6]). In the federated model, cognitive agents support clients and experts in problem solving tasks providing task automation and performance enhancement through augmented/amplified intelligence. Shared cognition resembles Lyytinen et al. 'federated innovation network' which has been

found to be conducive to innovation [24]. In this federated model instance, an intelligent system provides ML and AI enhancements to retrieve contextually meaningful data from the knowledge base (Fig. 3, Ref. [4]).

Building a federated CC system (Fig. 3, Ref. [4]) often begins by replicating and extending processes associated with the traditional internalized model through cognitive task automation. The learning process begins by capturing human knowledge and presenting it as a corpus, through ingestion into the knowledge-base (Fig. 3, Ref. [4]). The initial ML session, defined by ingesting relevant seed questions against the corpus, establishes a layer of ground truth within the cognitive model. The cognitive model is then refined in a design process mode through iterative sessions of unsupervised ML, supervised ML, and knowledge base harmonization. Model evaluation is based on performance testing before implementation through cognitive agents. Human cognitive processing has been described as a complementary combination of assimilation and accommodation processes. Through unsupervised ML, the CC system uses the ground truth to develop an initial knowledge base. The CC system recognizes inconsistency or ambiguity in the knowledge base, in other words, an instance of human accommodation, and orchestrates a supervised ML session to build that contextual-idiosyncrasy of knowledge into the system. From a knowledge engineering perspective, the system establishes a task context that reinforces the validity and appropriateness of declarative knowledge. Healthcare as a multi-domain translational science space is prone to variances of knowledge between domains. Knowledge harmonization resolves, through accommodation, discovered instances of inconsistency or ambiguity in the knowledge base within and between domains. With this knowledge layer based on ML, the CC system uses intelligent cognitive agents to interface more appropriately with the recipient (Refs. [5, 6]). Knowledge harmonization provides for multi-domain knowledge bases, robust longitudinal learning, and continuous improvement which supports the concept of eScience.

eScience

Based on deep machine learning, the system begins to identify and can apply alternatives with the greatest efficacies. In this manner, through extended time for learning and understanding, the system begins to build a basis for recognizing opportunities to recommend strategies in problem solving. Automation of executing recommended strategies brings the system to the threshold of eScience, where the computer iterates through problem solving cognitive tasks with which it is presented. Recurring acceptance and automation of a recommended problem-solving CC strategy in an obesity intervention program
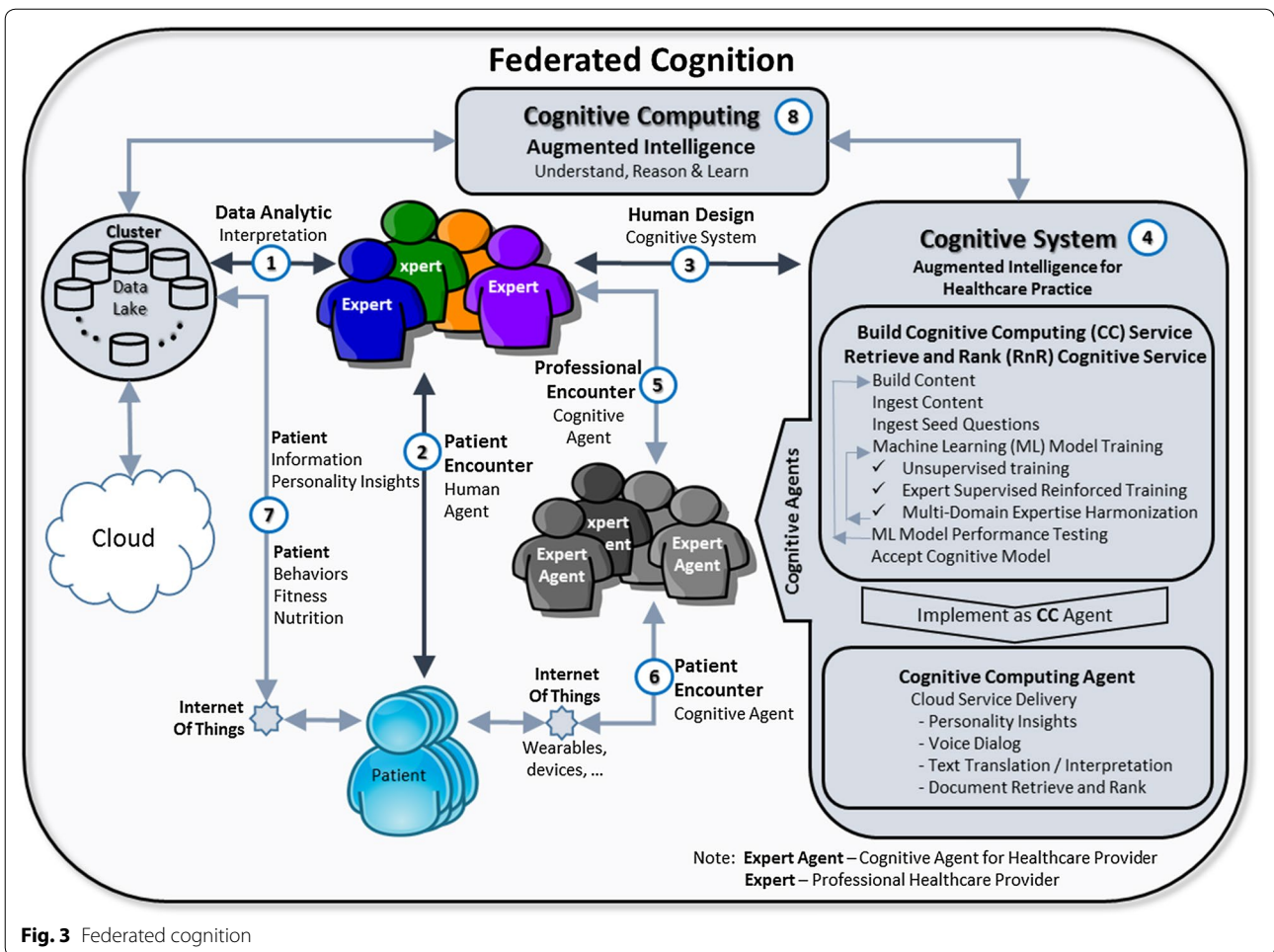
Marshall *et al. Health Inf Sci Syst* (2017) 5:13

Page 6 of 11

Fig. 3 Federated cognition

becomes a de facto data-driven *posterior* research methodology. In this case, the human collaborates with the computer by granting control of flexible strategies in problem solving to the CC based on semi-autonomous cognition (Fig. 4).

The continuum of semi-autonomous to autonomous cognition, from an eScience perspective, is a degree of human commitment to computer-controlled problem solving, including self-generated knowledge through deep learning, upon which to deliver cognitive services. Figure 5 presents autonomous cognition within an eScience model. Knowledge engineering would be the extent of direct human involvement in developing the service. Humans are not directly involved in the problem solving process as per problem type III from the Luconi et al. framework. Problem type III, expert systems, in the past were challenged to adopt over time without human knowledge engineering. Based on understanding and learning, CC systems evolve through experiences, and this newer capability for self-sustainability is a significant advancement over many older expert systems with static

knowledge bases. With a self-aware CC system, ambiguities and inconsistencies in the knowledge base are identified and rectified through supervised ML sessions as a knowledge engineering activity. Autonomous cognition is an eScience commitment to action based on CC system generated problem solving strategies. Autonomous cognition scales very well, is controllable, and can serve as a data-driven *posterior* research methodology supporting MOST healthcare initiatives.

Use case

In this use case, the intelligent system includes health, nutrition, and academic domains integrated into a knowledge base to support information retrieval. Each domain has its standardized nomenclature which is represented in the knowledge base along with an integrated metadata for a harmonized or unified nomenclature/knowledge-model combining the individual domains. Multiple domains within a support system increase the importance of having solid metadata definitions and standards that support user information requirements
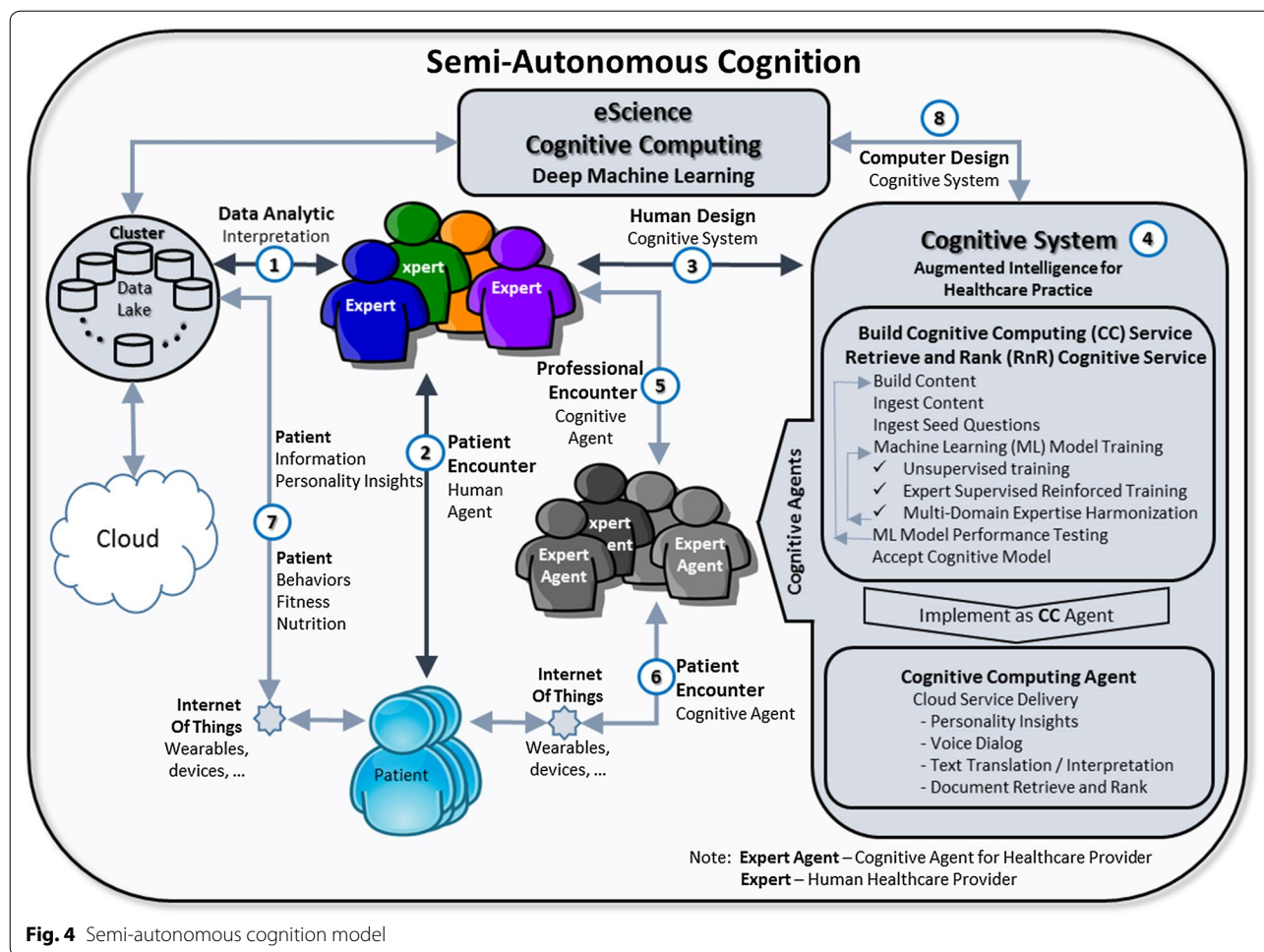
Marshall *et al. Health Inf Sci Syst* (2017) 5:13

Page 7 of 11

Fig. 4 Semi-autonomous cognition model

[25]. For each domain, a series of questions by domain are ingested into the CC system to develop an initial layer of ground truth. Process protocols considered best practices can also be stored as model abstractions within the knowledge base. It is essential to have a rich knowledge base that supports the variety of research perspectives associated with translational research [26].

The initial phase for health and nutrition domain management is to analyze the available data to develop baseline measures and cohorts of interest within each domain. Identification of the key factors that define baselines and sub-groups, followed by the initial design of recommended intervention programs, is a common protocol in the health and life sciences and will serve as an analytic base. In cases of interdisciplinary domains, the system uses unsupervised and reinforced (supervised) machine learning to harmonize the knowledge within and across domains. The content within the domains is then made available to the stakeholders through applications delivering intelligence based on the knowledge model. The knowledge presentation is predicated on the

defined role of the stakeholder(s) at hand in a given use-case scenario. The capabilities and process of knowledge integration across the domains support translational research and is a demonstration of the novel and innovative CC and eScience approaches to the study.

Health domain

Since physical fitness is a key tenet of health, applications assessing and tracking children's physical fitness are commonly integrated into physical education instruction. One existing health application that is universally used in public schools is called the FitnessGram®. Including a battery of assessments, the FitnessGram® developed by the Cooper Institute measures aerobic capacity, muscle fitness, and body composition as a means of minimizing health risk and educating parents about their child's health status. Commonly used in school settings to track and monitor children's health risk, the FitnessGram® was recently adopted by the Presidential Youth Fitness Program for use as national surveillance of health among school-aged children. Applications utilizing data such as
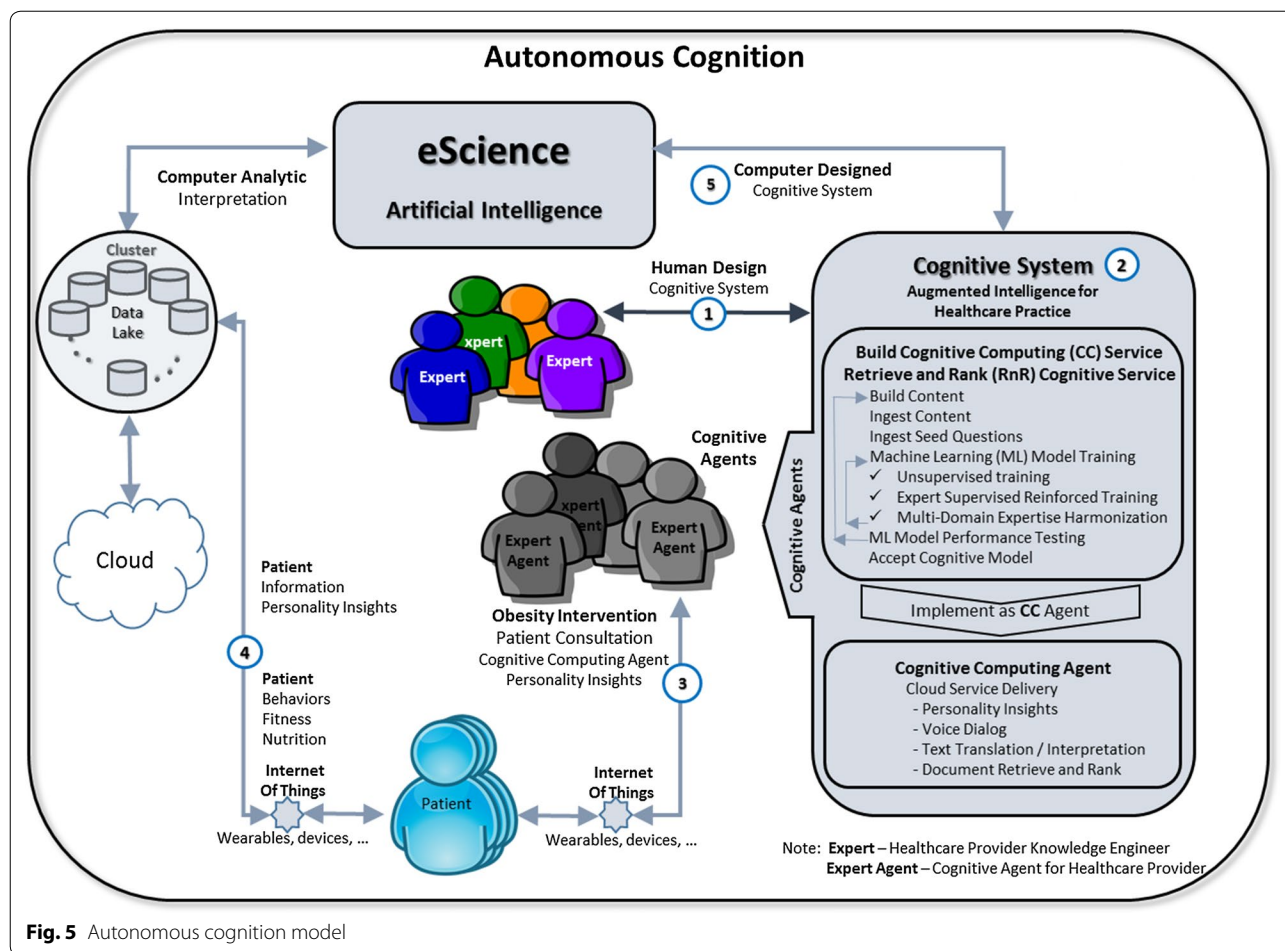
Marshall *et al. Health Inf Sci Syst (2017) 5:13*

Page 8 of 11

Fig. 5 Autonomous cognition model

the FitnessGram®, which focuses on the domain of physical fitness, are an integral way of addressing public health issues of childhood obesity, measured as health-related fitness.

Research suggests that physical fitness is associated with health outcomes such as diabetes and cardiovascular disease, as well as other adulthood morbidities [27–29]. Specifically, this application classifies child performance on a given assessment into the Health Fitness Zone, low health risk, and high health risk categories. Such an application, particularly given its potential to be integrated as national health surveillance, is an example of how CC and eScience intersects with disciplinary-specific applications already integrated in schools and physical education programming. Such data can contribute to the development and expansion of knowledge domains and intelligent cognitive agents that support obesity intervention programs. This has unique importance, given the contextual relevance of CC and eScience, the creation of disruptive catalysts, and its alignment with MOST interventions.

Multiphase optimization strategies (MOST) are aimed at identifying the components of health interventions that are the most effective [30]. Further, Strecher and colleagues would argue that there are some critical elements that must be in place before CC and eScience can be applied to health interventions: 1) employ externally valid methods of participant recruitment and 2) propose evidenced-based health intervention [30]. Only then can CC and eScience technologies be utilized to identify feasible program adaptations or extensions. Such CC and eScience technologies have already been applied to the examination of the effectiveness of self-regulation among smoking cessation programs [31]. Specifically, dynamical systems modeling was used to inform the smoking cessation intervention by focusing on the interactions between cravings, total cigarettes smoked in a given period, and self-regulation, thus paving the way for such CC and eScience strategies to be used as screening tools for all public health issues. The ability of deep machine learning algorithms to identify, at any point in time, which treatment may be the most appropriate for the individual is noteworthy [32].

Marshall *et al. Health Inf Sci Syst* (2017) 5:13

Page 9 of 11

Nutrition domain

The project is designed to capture nutrition data from the automated school cafeteria system. The cafeteria system includes the entire value-chain from back-end (kitchen) processes through to point of sale (POS) delivery. The POS data identifies cafeteria menu selections on an individual person basis. A student's purchase includes a selection of individual menu items. The automated data available from the POS includes an identified menu entrée with a number of side items. Nutritional values of the menu items are recorded in the back-end system which allows an extrapolation of nutritional value for an individual on a longitudinal basis and on a per event basis, i.e., purchase. Nutritional items captured in the system include calories, saturated fat, sodium, vitamin A, protein, cholesterol, vitamin C, carbohydrates, fiber, calcium, iron, total fat, ash, water, vitamin (IU), and Trans-fat. Nutrition knowledge includes descriptive and predictive analytic results associated with nutritional purchase behaviors of the subjects.

Academic domain

Academic performance will be standardized by using a structured coding system based on extensible markup language (XML). The XML standard, referred to as School Interoperability Framework (SIF), is used in more than 48 states, allows for more open data exchange, and enhances the computational nature of the data (SIF Wikipedia). The XML standard supports the integration of academic performance across school systems allowing aggregation and summarization of data. Academic performance is expected to be moderated by nutrition and health effects.

Design science

Cognitive computing and eScience models, as a collaborative human–machine effort, can support information discovery through design science, which is a discipline devoted to creating new data products. Following the design science model, data scientists create meaningful data products by defining processes of data acquisition, transformation, and integration where the data may come or be derived from a multitude of sources. Design science is a paradigm where" … knowledge and understanding of a problem domain and its solution are achieved in the building and application of the design artifacts." [33] As the authors state that design science will change information systems research methodologies, this paper proposes that design science is inherent in CC/eScience and both design science and CC/eScience will significantly impact health and life science research methodologies.

More recently, Roberts et al. propose design science as a human-centered research paradigm that provides a foundation for healthcare management, innovation, and practice [34]. The CC and eScience models in this study rely upon design science and implement design science in a continuous process improvement cycle by employing deep machine learning. Healthcare automation, as a design science use case, is presented in the next section to demonstrate the value of creating new data products to support desired CC and eScience processing capabilities.

Knowledge base development

With machine learning, ground truth represents the first layer of explicit knowledge entered into the CC/eScience system. Through unsupervised and supervised training sessions the knowledge base begins to develop and improve. Knowledge base development thus becomes a series of cyclical improvements through unsupervised and supervised training sessions (Fig. 6). Design science within the context of obesity intervention for children serves as the knowledge domain and, in essence defines the boundaries for the general corpus. The following iterative design process details the process of knowledge base development.

Obesity is associated with numerous co-morbidities and may be defined as an excessive amount of body fat [28]. Increases in obesity rates among children are considered by many to be a consequence of an individual's consumption of calorie-dense foods and level of physical activity. Intervention programs for children are often based on instituting changes in lifestyle that balance energy intake and energy expenditure [28]. Lifestyle changes have been noted to lower obesity among individuals; however, this weight loss benefit is often observed in the short-term but is longitudinally temporary. In this research, obesity intervention program design will emphasize lifestyle therapies (food consumption and activity) supported by technologies that develop personalized interfaces to enhance long-term program adherence [35].

Fig. 6 Knowledge base development

Marshall *et al. Health Inf Sci Syst (2017) 5:13*

Page 10 of 11
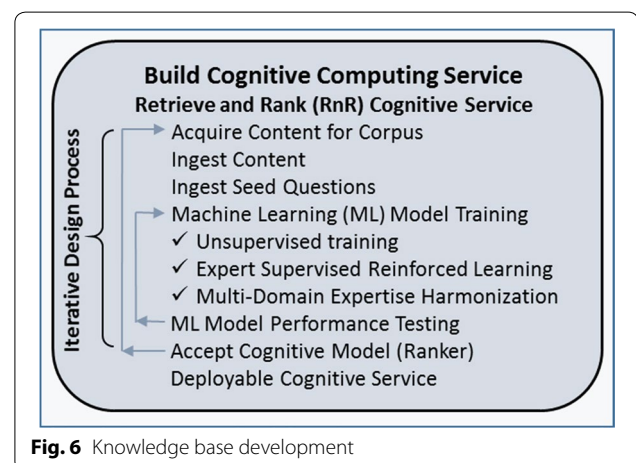
Given the system objective of developing obesity intervention programs, nutrition-related behaviors are important. A stronger analysis of subject nutrition-related behavior would be supported by additional data products related to menu entrée item selection. The current system as described lacks this analytic capability of entrée purchase analytics. With an objective of generating knowledge by transforming raw data into new data products, the following iterations of design science might be practiced.

A nutritionist interested in the distribution of food nutrition measures across individual students may wish to aggregate food item nutritional values by student into a scale (ranges 0–100) such as the Healthy Eating Index (HEI-2010) which assesses nutritional quality. In this process, data engineers would locate the relevant data, acquire that data, transform into the HEI, and create a visualization appropriate for the stakeholders to interpret students' HEI. Another iteration might involve acquiring a new data source of students' body mass index (BMI) and transforming the BMI data into a three-item classification representing less than moderate, moderate, and above moderate BMI. Following a similar process order, data engineers might combine the datasets, creating a cross-tabulation representing expected and actual groupings of HEI scale by BMI scale, providing insights to sub-population nutritional behaviors. With this analysis, the nutritionist can better plan meals with an interest in modifying student behavior. Also, the nutritionist can update the knowledge base to include the newly created data products and associated plans. The updated knowledge in the CC system can then use intelligent cognitive agents to personalize individual healthcare encounters for obesity intervention including HEI and BMI data.

Continuing the data design effort, an academic counselor might build upon the new data product by also including a scale representing student performance. With the given data, the academic counselor would be able to investigate potential relationships between HEI, BMI, and academic performance. Again, the counselor can update the knowledge base and, by doing so, enhance the behavior of cognitive agents. Another capability of the system is to use machine learning (ML) to develop solutions from the data lake. ML has the capability to create intelligence based on regression, classification, and/or recommender system models by combining dimensions from the nutrition, fitness, and academic domains to identify underlying explanations of the data. One use case example might be a director of operations, interested in increasing nutritional food selection and reducing inventory waste, using ML to recommend food item placement on the serving line to influence a greater selection of food-items such as fruits and vegetables. Through deep ML the food item placement model can be refined and improved based on empirical results. Integrating food item placement with other factors, such as fitness and academic performance, creates novel interpretations of what program components, dosages, and obesity intervention programs really work demonstrates how deep ML can support MOST research paradigms.

Cognitive agent training/reinforcement

The design processes just depicted demonstrate the ability of the system to support the integration of disparate data sources for intelligence creation, visualization, and use. The machine learning models in the CC component are then updated to create more-capable cognitive agents acting on the updated knowledge base. In this cognitive update, the system recognizes and works with the human experts to harmonize knowledge within and between domains to provide a richer encounter. Harmonization, using ML models, re-structures the knowledge base to provide information that is both non-ambiguous and highly relevant to the context. In an agent-based encounter, subject personality traits can be assessed over time and included in the cognitive agents to deliver more personalized intervention sessions. The system also takes an opportunistic approach to leveraging knowledge and includes reinforced learning to improve agent behavior. These capabilities highlight the dynamic capability of the system and its ability to support an evolving design science methodology. The newer research models can handle greater amounts of data, recognize patterns within the data, and deliver personalized motivational encounters that enhance individual and overall program success. The CC and eScience paradigms hold great potential to deliver effective MOST intervention programs that discover and implement what works [7].

Summary

The previous use cases demonstrate many important capabilities of CC and eScience research models. The models support the processing of large datasets allowing a relaxation of statistical modeling. Using classification, recommender, and intelligent content management systems on massive datasets provides a basis for creating and democratizing new knowledge through collective intelligence, ultimately resulting in more effective data-driven obesity intervention programs. Applications can be developed that contribute to the knowledge base and further the research objectives. Through cognitive agents, human experts augment and leverage their knowledge resulting in more effective and efficient healthcare delivery. Based on deep machine learning, the system leverages knowledge through iterative processes and supports continuous process improvement. In CC and eScience,

Marshall *et al. Health Inf Sci Syst* (2017) 5:13

Page 11 of 11

the cycle time of these iterations of knowledge accumulation to practice is highly compressed compared to more traditional approaches, and in many instances self-sustaining, that is with little or no researcher/human intervention. These factors make CC and eScience research paradigms appropriate for MOST intervention programs, especially those involving behavioral change.

Conclusion

This study, in comparing CC and eScience as research methodologies to a priori research has identified strengths and limitations of each methodological approach. This comparison has offered a validation of CC and eScience as new cornerstones for research and discovery in health and life sciences. In conclusion, CC and eScience do not negate hypothetico-deductive, i.e., a priori theory-based, research but instead implement and complement it in a more synergistic manner by augmenting and democratizing human intelligence. With the advent of health and life science big data, CC and eScience will facilitate intelligent research paradigms that, matched with more traditional research methods, promise to redefine a new era of data-driven discovery and practice. In this case, CC and eScience have a mutually beneficial disruptive impact on traditional hypothetico-deductive research methodologies in health and life sciences.

Author details

[1] Auburn University, Auburn, AL, USA. [2] School of Biomedical Informatics, Houston, TX, USA. [3] University of Texas, Austin, TX, USA. [4] The University of Texas Health Science Center at Houston, School of Public Health, Austin, TX, USA.

References

1. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2(1):3.
2. Hey T, Tansley S, Tolle KM, editors. The fourth paradigm: data intensive scientific discovery. Jim Grey, foreword, XVII, Microsoft Research; 2009.
3. Das R. Five technologies that will disrupt healthcare by 2020. Forbes.com, 30 Mar 2016.
4. Response to—request for information preparing for the future of artificial intelligence. Preparing for the future of artificial intelligence. http://research.ibm.com/cognitive-computing/ostp/rfi-response.shtml.
5. IBM. Cleveland clinic, IBM continue their collaboration to establish model for cognitive population health management and data-driven personalized healthcare, News release, Cleveland, OH and Armonk, NY. http://www-03.ibm.com/press/us/en/pressrelease/51290.wss. Accessed 22 Dec 2016
6. Garrison LP Jr. Universal health coverage—big thinking versus big data. J Int Soc Pharmaconomics Res. 2013;16(1):S1–3.
7. Norman GJ. Answering the "What Works?" question in health behavior change. Am J Prev Med. 2008;34(5):449.
8. Birch LL, Ventura AK. Preventing childhood obesity: what works? Int J Obes. 2009;33:S74–81.
9. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. Drug Alcohol Depend. 2007;88(Suppl 2):S31–40.
10. Deshpande S, Rivera DE, Younger JW, Nandola NN. A control systems engineering approach for adaptive behavioral interventions: illustration with a fibromyalgia intervention. Transl Behav Med. 2014;4(3):275–89.
11. Luconi FL, Malone TW, Scott-morton MS. expert systems: the next challenge for managers. Sloan Manag Rev. Summer. 1986;27(4):3.
12. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst. 2009;24:28.
13. Minvielle E, Waelli M, Sicotte C, Kimberly JR. Managing customization in health care: a framework derived from the services sector literature. Health Policy. 2014;117:216–27.
14. Sacks FM, Bray GA, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med. 2009;360:859–73.
15. Collins C. Murphy and stretcher: comparison of a phased experimental approach and a single randomized clinical trial for developing multicomponent behavioral interventions. Clin Trials. 2009;6(1):5–15.
16. Malin JL. Envisioning watson as a rapid-learning system for oncology. J Oncol Pract. 2013;9(3):155–7.
17. Wigner E. The unreasonable effectiveness of mathematics in the natural sciences. Commun Pure Appl Math. 1960;13(1):1–14.
18. Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. IEEE. 2009;24(2):8–12.
19. Klauser F, Albrechtslund A. From self-tracking to smart urban infrastructures: towards an interdisciplinary research agenda on Big Data. Surveill Soc. 2014;12(2):273–86.
20. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. Big Data. 2013;1(2):85–99.
21. Newell A. Reasoning: problem solving and decision processes: the problem space as a fundamental category. In: Nickerson R, editor. Attention and performance VIII. Hillsdale: Erlbaum; 1980.
22. Newell A, Simon HA. Human problem solving. Englewood Cliffs: Prentice-Hall; 1972.
23. Frey CB, Osborne MA. How susceptible are jobs to computerization? Technol Forecast Soc Change. 2017;2017(114):254–80.
24. Lyytinen K, Yoo Y, Boland RJ Jr. Digital product innovation within four classes of innovation networks. J Inf Syst. 2016;26(1):47–75.
25. Berman JJ. Principles of big data: preparing, sharing and analyzing complex information. Amsterdam: Elsevier, Inc; 2013.
26. Sadasivam RS, Tanik MM. A meta-composite software development approach for translational research. J Med Syst. 2013;37(3):9935.
27. McGuire S. Accelerating progress in obesity prevention: solving the weight of the nation, *Advanced Nutrition* 2012 1:3 (5) 7808-709. Institute of Medicine (IOM). Washington, DC: The National Academies Press; 2012.
28. Atay Z, Bereket A. Current status on obesity in childhood and adolescence: prevalence, etiology, co-morbidities and management. Obes Med. 2016;3:1–9.
29. Hekler EB, Buman MP, Poothakandiyil N, Rivera DE, Dzierzewski JM, Morgan AA, et al. Exploring behavioral markers of long-term physical activity maintenance: a case study of system identification modeling within a behavioral intervention. Health Educ Behav. 2013;40(10):51S–62S.
30. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res. 2008;10(5):e36.
31. Timms KP, Rivera DE, Piper ME, Collins LM. A hybrid model predictive control strategy for optimizing a smoking cessation intervention. In: Proceedings of the 2014 American control conference; 2014. p. 2389–2394.
32. McClure AC, Stoolmiller M, Tanski SE, Engels RC, Sargent JD. Alcohol marketing receptivity, marketing-specific cognitions, and underage binge drinking. Alcohol Clin Exp Res. 2012;. doi:10.1111/j.1530-0277.2012.01932.x.
33. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. Manag Inf Syst Q. 2004;28(1):75–105.
34. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc J Deliv Sci Innov. 2016;4(1):11–4.
35. Dieris B, Reinehr T. Treatment programs in overweight and obese children: How to achieve lifestyle changes? Obes Med. 2016;3:10–6.