

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 April 2009 (16.04.2009)

PCT

(10) International Publication Number
WO 2009/049278 A1

(51) International Patent Classification:
G06Q 10/00 (2006.01)

(74) Agents: LOREN, Ralph, A. et al.; Edwards Angell Palmer & Dodge LLP, P.O. Box 55874, Boston, MA 02205 (US).

(21) International Application Number:
PCT/US2008/079674

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date: 12 October 2008 (12.10.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/998,768 12 October 2007 (12.10.2007) US
60/998,669 12 October 2007 (12.10.2007) US
61/070,067 20 March 2008 (20.03.2008) US

(71) Applicant (for all designated States except US): PATIENTSLIKEME, INC. [US/US]; 222 Third Street, Suite 0234, Cambridge, MA 02142 (US).

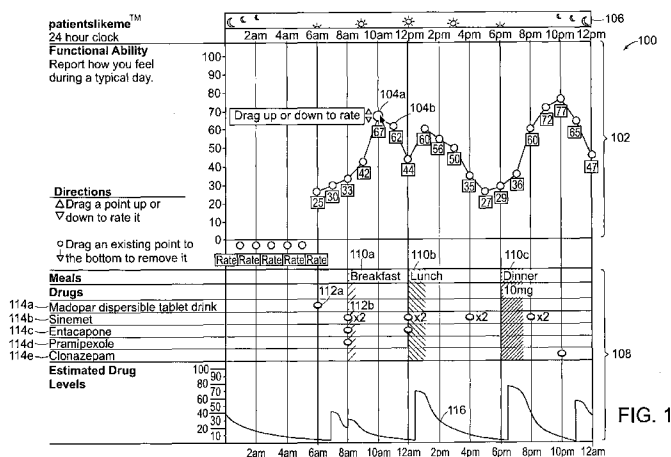
(72) Inventors; and

(75) Inventors/Applicants (for US only): HEYWOOD, James [US/US]; 222 Third Street, Suite 0234, Cambridge, MA 02142 (US). HEYWOOD, Benjamin [US/US]; 79 Antrim Street, Cambridge, MA 02139 (US).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

(54) Title: SELF-IMPROVING METHOD OF USING ONLINE COMMUNITIES TO PREDICT HEALTH-RELATED OUTCOMES



(57) Abstract: The invention is directed, in part, to method of using self-reported health data in online communities to predict significant health events in life-changing illnesses to improve the lives of individuals and to improve patient self-management. The invention provides a method for providing real-time personalized medical predictions for an individual patient. The method includes: providing a database containing patient information for a plurality of other patients including one or more attributes for each patient in the database; constructing a model of a disease based on disease progressions for the plurality of patients; receiving a request from the individual patient, the patient associated with one or more attributes; and making a real-time prediction for the individual patient based on the mode and the individual patient's attributes.

SELF-IMPROVING METHOD OF USING ONLINE COMMUNITIES TO PREDICT HEALTH-RELATED OUTCOMES

Related Application

5 This application claims priority to U.S. Provisional Patent Application No. 60/998,669, filed on October 12, 2007, U.S. Provisional Patent Application No. 60/998,768, filed on October 12, 2007, and U.S. Provisional Patent Application No. 61/070,067, filed on March 20, 2008. The entire contents of each of these applications is hereby incorporated by reference herein.

10

Technical Field

The present invention is directed to a method of using self-reported health data in online communities to predict significant health events in life-changing illnesses to improve the lives of individuals and to improve patient self-management.

15

Background of the Invention

According to the World Health Organization, chronic diseases are now the major cause of disability and death worldwide, accounting for 59% of 57 million deaths annually and 46% of the global burden of disease. According to the U.S. Centers for Disease Control and Prevention, more than 90 million Americans live with chronic illnesses, accounting for more than 75% of the national \$1.4 trillion bill for medical care costs. Chronic diseases also account for one-third of the years of potential life lost before age 65. Although widespread illnesses such as cardiovascular disease or diabetes are well-characterized in terms of risk-factors, prevention, and treatment, there are a host of under-researched and untreatable conditions; the National Organization for Rare Disorders (NORD) tracks approximately 6,000 rare disorders which, altogether, affect 23 million Americans.

Research into the relationship between health behavior and outcomes suffers from a variety of methodological flaws. There is often insufficient funding for prospective follow-up studies, service provision may be dependent on research staff, sample sizes are small, and particularly in the case of rare diseases, recruitment is difficult. Furthermore, results from clinical research tend to be written in scientific jargon, are difficult for the general public to understand, and refer to group averages rather than individual outcomes.

30

Accordingly, there is a need for an effective process to (i) collect data on interventions and health outcomes, (ii) model the likely course of a disease for an individual on the basis of their background and experience to-date, (iii) provide information on likely outcomes to the individual to help them manage their condition,
5 and (iv) improve the model to improve the accuracy of predictions made.

The term “intervention” refers any event that has a positive, negative, or neutral effect on one or more medical conditions. The term intervention includes a variety of activities including, but not limited to, administration of a medication, administration of a remedy, administration of a nutritional supplement, administration
10 of a vitamin, exercise, physical therapy, massage, stretching, consumption of food, rest, and sleep.

Summary of the Invention

The present invention meets the foregoing need and provides an effective
15 method of predicting health outcomes for an individual with a life-changing health condition, which will result in greater empowerment over their healthcare and better outcomes apparent from the discussion herein.

The invention provides a method for providing real-time personalized medical predictions for an individual patient. The method includes: providing a database
20 containing patient information for a plurality of other patients including one or more attributes for each patient in the database; constructing a model of a disease based on disease progressions for the plurality of patients; receiving a request from the individual patient, the patient associated with one or more attributes; and making a real-time prediction for the individual patient based on the mode and the individual
25 patient’s attributes.

The one or more attributes can include at least one selected from the group consisting of: age, race, ethnicity, gender, height, weight, body mass index (BMI), body volume index (BVI), genotype, phenotype, severity of the disease, progression rate of the disease, measures of functional ability, quality of life, interventions, and
30 remedies.

The disease can include at least one selected from the group consisting of: neurological diseases, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson’s Disease, Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer blood cancer,

fibromyalgia, epilepsy, post traumatic stress disorder, traumatic brain injury, cardiovascular disease, osteoporosis, chronic obstructive pulmonary disease, arthritis, allergies, autoimmune diseases, and lupus.

5 The data returned can include individual data for one or more members of the set of other patients. The data returned can include aggregate data for one or more members of the set of other patients. The method can include processing a request from the patient to view individual data.

10 The model can be based on data for a subset of the plurality of patients. The method can include processing a request from the patient to modify a composition of the subset of the plurality of patients. The composition of the subset of other patients can defined by fuzzy logic. The step of modifying the composition of the subset of the plurality of patients can include modifying the range of attributes of patients within the subset. The step of modifying the composition of the subset of the plurality of the patients can include modifying the importance of attributes of patients in
15 composing the subset. The method can include conducting a multivariate pattern matching search of data related to the plurality of patients.

 The method can include calculating a confidence interval for the prediction. The step of calculating a confidence interval for the prediction can include: selecting a set of reported data points from the plurality of other patients, for each of the reported
20 data points in the set: obtaining a data set for the corresponding other patient to the reported data point calculating a predicted value with the data set and the model, and calculating an error between the predicted value and the reported data point; producing a distribution of the errors; and calculating a confidence interval from the distribution.

25 The set of reported data points can include n closest reported data points to the prediction. The set of reported data points can include reported data points within an ellipsoid defined by a distance metric. The size of the data set for the corresponding other patient can be comparable to a quantity of attributes associated with the individual patient.

30 The method can include analyzing an effect of an intervention by measuring a difference between a prediction absent the intervention and a reported outcome with the intervention. The difference can be measured for a plurality of individual patients. The difference can be compared to the distribution of error. The difference can be compared to the confidence interval for the model.

The method can include identifying one or more of the differences that exceed the confidence interval for the model. The method can include assembling a distribution of the differences for the plurality of individual patients, and computing a standard error for the distribution. The confidence interval can be calculated with a
5 chi-square test. The confidence interval can be calculated from a measure of variance of the individual patient's attributes. The confidence interval can be calculated by comparing the individual patient's attributes to a model fit for the individual patient using the model.

The invention also provides a computer-readable medium whose contents
10 cause a computer to perform a method for providing real-time personalized medical predictions for an individual patient. The method includes: providing a database containing patient information for a plurality of other patients including one or more attributes for each patient in the database; constructing a model of a disease based on disease progressions for the plurality of patients; receiving a request from the
15 individual patient, the patient associated with one or more attributes; and making a real-time prediction for the individual patient based on the mode and the individual patient's attributes.

The invention also provides a method for providing real-time personalized medical predictions. The method includes: gathering patient-submitted information
20 from a community of patients having a disease, the information including medical condition metrics and intervention data; utilizing the patient-submitted information to form a model of the disease; and predicting the progression of the disease in a particular patient by applying information submitted by an individual patient to the model.

25 The information submitted by the individual patient can include a date of onset of the disease. The step of predicting the progression of the disease can be performed in real time. The step of predicting the progression of the disease can include providing a confidence interval. The step of predicting the progression of the disease can include providing a graphical prediction. The graphical prediction can be a line
30 chart depicting development of the disease with regard to a rating scale. The disease can include at least one selected from the group consisting of: neurological diseases, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson's Disease, Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer blood cancer, fibromyalgia, epilepsy,

post traumatic stress disorder, traumatic brain injury, cardiovascular disease, osteoporosis, chronic obstructive pulmonary disease, arthritis, allergies, autoimmune diseases, and lupus. The model can be a model of the disease's pathology.

The invention provides a method for providing real-time personalized medical
5 predictions for an individual patient. The method includes providing a database containing patient information for a plurality of patients including one or more attributes for each patient in the database; providing a graphical user interface displaying one or more attributes of the individual patient, the graphical user interface allowing the patient to formulate a search request specifying at least one of the
10 attributes; searching the database of patient information for patients having the specified one or more attributes; providing a model of a disease based on disease progressions for the patients having the specified one or more attributes; and making a real-time prediction for the individual patient based on the model.

The one or more attributes can include at least one selected from the group
15 consisting of: age, race, ethnicity, gender, height, weight, body mass index (BMI), body volume index (BVI), genotype, phenotype, severity of the disease, progression rate of the disease, measures of functional ability, quality of life, interventions, and remedies.

The disease can include at least one selected from the group consisting of:
20 neurological diseases, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson's Disease, Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer blood cancer, fibromyalgia, epilepsy, post traumatic stress disorder, traumatic brain injury, cardiovascular disease, osteoporosis, chronic obstructive pulmonary disease, arthritis,
25 allergies, autoimmune diseases, and lupus.

The data returned can include individual data for one or more members of the set of other patients or aggregate data for one or more members of the set of other patients. The method can include processing a request from the patient to view individual data. The method can also include processing a request from the patient to
30 modify a composition of the set of other patients. The composition of the set of other patients can be defined by fuzzy logic.

The step of modifying the composition of the set of other patients can include modifying the range of attributes of patients within the set. The step of modifying the composition of the set of other patients can include modifying the importance of

attributes of patients in composing the set. The method can include conducting a multivariate pattern matching search of data related to the other patients.

The outcome data can include at least one medical condition selected from the group consisting of: occurrence of epilepsy, occurrence of migraine, pain, fatigue,
5 cognitive ability, anxiety, mobility, dexterity, and occurrence of allergies.

The method can include calculating a confidence interval for the model.
The method can also include analyzing an effect of an intervention by measuring a difference between an expected outcome absent the intervention as predicted by the model and a reported outcome with the intervention. The difference can be measured
10 for a plurality of individual patients. The difference can be a sum of the observational error rate based on a quality of the plurality of patient's pre-model data and a variation from the model. The difference can be compared to the confidence interval for the model.

The invention also provides a computer-readable medium whose contents
15 cause a computer to perform a method for providing real-time personalized medical predictions for an individual patient. The method includes: providing a database containing patient information for a plurality of patients including one or more attributes for each patient in the database; providing a graphical user interface displaying one or more attributes of the individual patient, the graphical user interface
20 allowing the patient to formulate a search request specifying at least one of the attributes; searching the database of patient information for patients having the specified one or more attributes; providing a model of a disease based on disease progressions for the patients having the specified one or more attributes; and making a real-time prediction for the individual patient based on the model.

25 The invention also provides a method for providing real-time personalized medical predictions. The method includes: gathering patient-submitted information from a community of patients having a disease, the information including medical condition metrics and intervention data; utilizing the patient-submitted information to form a model of the disease; and predicting the progression of the disease in a
30 particular patient based on information submitted by an individual patient.

The step of predicting the progression of the disease can be performed in real time. The step of predicting the progression of the disease can include providing a confidence interval. The step of predicting the progression of the disease can include

providing a graphical prediction. The graphical prediction can be a line chart depicting development of the disease with regard to a rating scale.

The disease can include at least one selected from the group consisting of: neurological diseases, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS),
5 Parkinson's Disease, Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer blood cancer, fibromyalgia, epilepsy, post traumatic stress disorder, traumatic brain injury, cardiovascular disease, osteoporosis, chronic obstructive pulmonary disease, arthritis, allergies, autoimmune diseases, and lupus.

10 The invention also provides a method for providing personalized medical information comprising: providing a database containing patient information for a plurality of patients including one or more attributes for each patient in the database; providing a graphical user interface displaying one or more attributes of a patient, the graphical user interface allowing the patient to formulate a search request specifying
15 at least one of the attributes; searching the database of patient information for patients having the specified one or more attributes; and returning data to the patient identifying a set of other patients having the specified one or more attributes.

The one or more attributes can include at least one selected from the group consisting of: age, race, ethnicity, gender, height, weight, body mass index (BMI),
20 body volume index (BVI), genotype, phenotype, disease, disease severity, disease progression rate, measures of functional ability, quality of life, interventions, and remedies.

The database can include one or more correlations between an attribute and at least one secondary attribute selected from the group consisting of: quality of life,
25 functional ability, pain, and treatment intensity.

The disease can include at least one selected from the group consisting of: Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson's Disease, Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer blood cancer, fibromyalgia, epilepsy,
30 post traumatic stress disorder, and traumatic brain injury.

The data returned can include individual data for one or more members of the set of other patients. The data returned can include aggregate data for one or more members of the set of other patients.

The method can include processing a request from the patient to view individual data. The method can also include processing a request from the patient to modify a composition of the set of other patients. The composition of the set of other patients can be defined by fuzzy logic. Modifying the composition of the set of other patients can include modifying the range of attributes of patients within the set. Modifying the composition of the set of other patients can include modifying the importance of attributes of patients in composing the set. The composition of the set of other patients can be defined by an optimal matching algorithm on a graph of attribute similarity metrics. The composition of the set of other patients can be defined by a scalar-vector decomposition on a matrix of similarities of attributes of the set of other patients. The method can also include conducting a multivariate pattern matching search of data related to the other patients.

The invention also provides a computer-readable medium whose contents cause a computer to perform a method for providing personalized medical information. The method includes the steps of: providing a database containing patient information for a plurality of patients including one or more attributes for each patient in the database; providing a graphical user interface displaying one or more attributes of a patient, the graphical user interface allowing the patient to formulate a search request specifying at least one of the attributes; searching the database of patient information for patients having the specified one or more attributes; and returning data to the patient identifying a set of other patients having the specified one or more attributes.

The invention also provides a method for providing personalized medical information. The method includes the steps of: providing a database containing patient information for a plurality of patients including one or more attributes for each patient in the database; providing a graphical user interface displaying one or more attributes of a patient, the graphical user interface allowing the patient to formulate a search request specifying at least one of the attributes; searching the database of patient information for patients having the specified one or more attributes; and providing outcome data for other patients that previously had similar attributes to the specified one or more attributes.

The one or more attributes can include at least one selected from the group consisting of: age, race, ethnicity, gender, height, weight, body mass index (BMI), body volume index (BVI), genotype, phenotype, disease, disease severity, disease

progression rate, measures of functional ability, quality of life, interventions, and remedies.

The disease can include at least one selected from the group consisting of: Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson's Disease,
5 Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer, blood cancer, fibromyalgia, epilepsy, post traumatic stress disorder, and traumatic brain injury.

The data returned can include individual data for one or more members of the set of other patients. The data returned can include aggregate data for one or more
10 members of the set of other patients.

The method can also include processing a request from the patient to view individual data. The method can also include processing a request from the patient to modify a composition of the set of other patients. The composition of the set of other patients can be defined by fuzzy logic. Modifying the composition of the set of other
15 patients can include modifying the range of attributes of patients within the set. Modifying the composition of the set of other patients can also include modifying the importance of attributes of patients in composing the set.

The method can also include conducting a multivariate pattern matching search of data related to the other patients. The outcome data can include at least one
20 medical condition selected from the group consisting of: occurrence of epilepsy, occurrence of migraine, pain, fatigue, cognitive ability, anxiety, mobility, dexterity, and occurrence of allergies.

The invention also provides a computer-readable medium whose contents cause a computer to perform a method for providing personalized medical
25 information. The method can include: providing a database containing patient information for a plurality of patients including one or more attributes for each patient in the database; providing a graphical user interface displaying one or more attributes of a patient, the graphical user interface allowing the patient to formulate a search request specifying at least one of the attributes; searching the database of patient
30 information for patients having the specified one or more attributes; and providing outcome data for other patients that previously had similar attributes to the specified one or more attributes.

Accordingly, a method of modelling an individual's disease progression using an online community can include the steps of creating an online community for

people with life-changing illnesses; creating membership accounts for patients joining the online community; receiving personal and medical information from the patient; categorizing and storing the received personal and medical information to a member database; modelling the relationship between an individual's background and current
5 health status in comparison to data received from other patients like them; making predictions about future outcomes which are presented to the patient with levels of confidence; allowing predictions to be modified on the basis of received information from health providers or administrative records systems, such as utilization of health services, laboratory test results, diagnostic procedures, therapeutic procedures, or
10 from other measurement systems of phenotypic or genotypic characteristics; and receiving feedback from members over the course of time to validate or modify the model accordingly in order to improve the model.

The present invention also provides a method of displaying visual information to a patient about their likely predicted disease states, health events, and health
15 outcomes which includes estimations of confidence surrounding the likelihood of such outcomes, which can change over time in response to more data from themselves, from other sources such as health providers or administrative systems, or other members of the site.

The invention also provides a method of grouping patients by background and
20 illness-specific characteristics including but not limited to: the presence or absence of a genetic mutation; the presence or absence of a genetic polymorphism; the presence or absence of a pattern of familial inheritance apparent from a family history; the presence or absence of a known or unknown proteome sequence; the staging of disease progression according to self-report or assessment by a healthcare
25 professional; functional outcome assessed by a self-report questionnaire or by a healthcare professional; demographic information submitted by the patient including, but not restricted to age, sex, ethnicity, socio-economic status, health behaviours, diet, exercise, smoking history, drug use, surgical history, past and present geographical location, personality, and/or chance; and the rate, nature, direction of change, or
30 interaction of any of the above. The method of grouping can include but is not limited to multiple regression to identify significant predictor variables such as suggested above.

The present invention also provides a method by which the grouping of patients permits the construction of a mathematical model establishing the likelihood

of important medical outcomes such as, without limitation, when a patient may need a health intervention (e.g., surgery, tablet, therapy, assistive technology, home modification, nutritional supplement, lifestyle change and the like); the likelihood of adverse events in response to intervention (e.g., side effects, injuries, death and the like); the likelihood that the patient will develop a symptom, condition, or disease; the likelihood that a child or other relative of the patient will develop a symptom, condition, or disease; the likelihood that a patient will need to receive care in an institution as opposed to receiving care at home; the point at which the costs of paying for medical care will achieve a certain level; and the degree of improvement which might be experienced if the patient chooses to start a given intervention.

The present invention also includes a method by which a predicted event is displayed on the patient's medical profile at a future date, such as the events as noted above. The predicted event may also be presented with a level of confidence dependent on the quantity and quality of data entered by the user or by data available from other sources and users of the system.

The present invention also includes a method by which models used to predict future events are strengthened or weakened by ongoing feedback from users once the predicted event has come to pass.

The present invention also provides a server for facilitating a Web site portal that collects and analyzes information related to patients having at least one common characteristic such as a disorder, wherein the server communicates with clients via a distributed computing network and the patients and related caregivers can access the Web site portal via a client, and wherein the server comprises: (a) a memory storing an instruction set and historical data related to a plurality of patients; and (b) a processor for running the instruction set, the processor being in communication with the memory and the distributed computing network, wherein the processor is operative to: (i) receive additional data related to the patients and add the additional data to the historical data; (ii) model the historical data generally for a subset of the patients with a second common characteristic; (iii) model the historical data for an individual patient within the subset of the patients; (iv) make a prediction of a timeframe for a future event based on the modelled subset and individual patient historical data; (v) add the prediction to the historical data; (vi) analyze the prediction based on the historical data to determine a confidence parameter; and (vii) provide a display of the historical data modelled for the subset, the historical data modelled for

the individual patient, the prediction for the individual patient, and the confidence parameter.

It should be appreciated that the present invention can be implemented and utilized in numerous ways, including without limitation as a process, an apparatus, a system, a device, a method for applications now known and later developed or a computer readable medium. These and other unique features of the system disclosed herein are readily apparent from the subject description and the accompanying disclosure and drawings.

Brief Description of the Drawings

For a fuller understanding of the nature and desired objects of the present invention, reference is made to the following detailed description taken in conjunction with the accompanying drawing figures wherein:

FIG. 1 is a diagram depicting an exemplary graphical element.

FIG. 2 is a diagram depicting the mapping of the ALSFRS scale to specific spinal nerves.

FIG. 3 is a diagram depicting an exemplary user interface for viewing and refining a group of similar patients and prediction of disease progression.

FIG. 4 is a diagram depicting a method of producing a confidence interval for a predictive model.

FIG. 5 is a diagram depicting an exemplary graphical element displaying a prediction and a confidence interval.

FIG. 6 is a diagram depicting the construction of a pathological model of ALS from the ALSFRS questionnaire, the progression of the pathological model, and the use of the pathological model to predict answers on a future ALSFRS questionnaire.

Definitions

The instant invention is most clearly understood with reference to the following definitions:

As used in the specification and claims, the singular form “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise.

The term “disease” refers to an abnormal condition of an organism that impairs bodily functions. The term disease includes a variety of physical ailments including, but not limited to, neurological diseases (*e.g.*, Amyotrophic Lateral

Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson's Disease), Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), cancers (*e.g.*, bladder cancer, blood cancer, breast cancer, colorectal cancer, endometrial cancer, leukemia, lung cancer, lymphoma, ovarian cancer, pancreatic cancer, prostate cancer, and skin cancer), diabetes, digestive disorders (*e.g.*, irritable bower syndrome, gastro esophageal reflux disease, and Crohn's Disease), cardiovascular diseases, osteoporosis, chronic obstructive pulmonary disease (COPD), arthritis, allergies, geriatric diseases, and autoimmune diseases (*e.g.*, lupus). The term disease also include mental ailments including, but not limited to, depression, anxiety disorders, post traumatic stress disorder, mood disorders, psychotic disorders, personality disorders, and eating disorders.

The term "medical condition" refers to a manifestation of a disease such as a symptom. For example, if a patient suffers from Amyotrophic Lateral Sclerosis (ALS), the patient may experience one or more medical conditions such as dysphagia (impaired swallowing).

Detailed Description of the Invention

The invention is directed, in part, to method of using self-reported health data in online communities to predict significant health events in life-changing illnesses to improve the lives of individuals and to improve patient self-management.

Data Acquisition

Self-reported health data can be gathered from a number of sources such as the PatientsLikeMe™ service available at www.patientslikeme.com.

An online community can be created to allow patients to contribute information about themselves, their diseases, their medical conditions, and their interventions. Each patient can register for one or more communities focused on a particular disease. For example, a patient can join an ALS community. As part the registration process, the patient can enter various demographic and/or medical information. Exemplary information can include: age, race, ethnicity, gender, height, weight, body mass index (BMI), body volume index (BVI), genotype, phenotype, disease, disease severity, disease progression rate, measures of functional ability, quality of life, interventions, remedies, and medical data such as tests. Patient information can also include historical or environmental data such as weather data for the patient's environment (*e.g.*, temperature, humidity, pollen count, air quality) and

the patient's past exposure to the sun. Patient information can also include personality information. Personality information can be represented by various classification systems such as the DISC assessment, Enneagram of Personality, Keirsey Temperament Theory, and the Meyer-Briggs Type Indicator. Genotype can be
5 determined through known SNP (single nucleotide polymorphism) or full-genome sequencing techniques.

After registration, the patient periodically inputs information about one or more medical conditions and one or more remedies. For example, an ALS patient can indicate when she sleeps, eats, and takes various medications such as riluzole.
10 Likewise, the ALS patient can enter data on their functional ability at various times throughout the day. Rating scales for assessing ALS patients include the Appel ALS rating scale and the ALS Functional Rating Scale.

Referring to FIG. 1, user interface 100 includes a medical condition metric portion 102, which allows the patient to input a medical condition metric (in this
15 example, the patient's functional ability). The user can place multiple data points 104 in the medical condition metric chart, which includes a time scale. Data points 104 can be adjusted with respect to time and/or magnitude. For example, if the patient is indicating how she feels now or at a designated time, the patient can be limited to moving data point 104 up or down. Alternatively, the patient can input data for a time
20 by dragging the data point to the left or right. The patient can be restricted in some embodiments from setting a data point in the future.

User interface 100 also includes an intervention portion 108. Intervention portion 108 allows the patient to record one or more interventions such as administration of a medication, administration of a remedy, administration of a
25 nutritional supplement, administration of a vitamin, exercise, physical therapy, massage, stretching, consumption of food, rest, and sleep. For example, the patient can designate when meals are eaten by adjusting bars 110a, 110b, and 110c to indicate the beginning and ending of the meal. Likewise, the patient can indicate when one or more drugs 114a-114e are administered by placing markers 112 (which may depict
30 pills) on a time scale.

Various types of remedies can be scheduled for specific times. For example, the patient can be prescribed to take madopar at 6 A.M. In this situation, user interface 100 can display a medication schedule. The patient can modify this schedule to reflect the actual administration by dragging marker 112a. Likewise,

the patient can indicate that the drug was consumed by clicking on the marker 112a. Clicking on the marker can change the appearance of the marker 112a (*e.g.*, its color) and thus can be used by patients, particularly patients with memory problems, to more faithfully follow a medication program.

5 User interface 100 can also include pharmacokinetic data, such a pharmacokinetic curve 116 that depicts the concentration of a medication within the patient over time. Multiple pharmacokinetic curves 116 can be depicted in various colors or patterns to reflect varying pharmacokinetic properties of various medications.

 This patient information is then stored in various formats. The data can be
10 stored in a relational database. Suitable relational databases includes DB2® and INFORMIX® both available from IBM Corp. of Armonk, New York; MICROSOFT JET® and MICROSOFT SQL SERVER® both available from the Microsoft Corp. of Redmond, Washington; MYSQL® available from the MySQL Ltd. Co. of Stockholm, Sweden; ORACLE® Database, available from Oracle Int'l Corp of
15 Redwood City, California; and SYBASE® available from Sybase, Inc. of Dublin, California. Additionally or alternatively, the data can be stored in a state model.

A Priori Disease Modelling Algorithms

 The invention also provides an *a priori* method for modelling diseases. This method can include an *a priori* model of a progressive disease. This method does not
20 rely on an underlying description of a pathology, and thus may be used even for diseases for which a pathology is unknown.

 The model allows for the prediction of the state of a disease at some point in time (*e.g.*, past, present, and future) based on data provided by an individual patient. The data can be any of the items discussed herein, including, but not limited to
25 medical condition metrics (*e.g.* rating scales) and the date of onset of the disease. The model is designed to provide an accurate and reliable prediction even where the data provided by the individual patient and the community of patients is heterogeneous (*e.g.*, reported at varying time intervals).

 This model describes the state of a disease for a patient at any point in time as
30 a scalar-valued or vector-valued function. For example, in a model of ALS, the state can be an integer value in the closed interval between 0 and 48 (the range of the overall Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R) scale) or the state can be a twelve-dimensional vector indicating the responses to all

twelve ALSFRS-R questions. The ALSFRS-R is a 48-point scale including an questions assessing the patient ability to walk, breath, communicate, etc.

The progression of a disease can be described by the time required since some event (*e.g.* onset or diagnosis of the disease) to reach some value in the above-described disease state function. It may also be described as a differential equation or as the time required to cross some boundary in the state space. This progression can be held constant for any particular patient (and thus be treated as an inherent attribute of the patient's disease) or it can be variable over the course of the disease.

Using the progression rate as one dimension, a multivariate function may describe set of all patients in a population. The population can include a group of patients, for example, all patients participating a community such as those provided through the PatientsLikeMe™ service. Each patient is associated with a set of data points describing their disease state at different points in time. In the above example of a scalar-valued disease state function (overall ALSFRS-R score), the function would exist as a surface (scalar function) of two dimensions: (i) progression rate and (ii) time since some event (*e.g.* onset of the disease). In the example of a vector-valued disease state function, the function would be a twelve-dimensional vector field as a function of two dimensions ($f(R^2) \rightarrow R^{12}$).

An expression of this model function can be determined using numeric methods such as the Levenberg-Marquardt algorithm to provide an optimal fit according to some appropriate choice of function. The Levenberg-Marquardt algorithm is described in publications such as Donald Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, 11 SIAM Journal on Applied Mathematics 431-41 (1963). For example, a scalar bi-cubic function of time and progression rate can provide a good functional form to describe a change in overall ALSFRS-R score over time.

The fitness of the model may be improved through an iterative procedure whereby several first-generation parameters are replaced with second-generation variants. For example, the initial value of progression rate can be replaced for each patient in the data set with a value that provides a least-mean-squares fit to the model. Thus, the initial approximation of progression rate is improved by iteration. The model function can then be re-fit to the new set of parameters and this process can be repeated one or more times, until some stopping conditions (*e.g.* convergence or lack of further change in the progression rates) are met.

A set of error bounds, or a confidence interval, can be determined from the actual error rate of the model. To determine the error bounds around a particular prediction (*i.e.*, a prediction that a specific patient will have a certain disease state at a specified time), the method can consider all known patient-reported data points within
5 a certain radius of the prediction, or alternatively, the n nearest patient-reported data points for some positive integer n . In either case, a valid definition of radius (distance metric) can be derived from the patient's progression rate or the derivative of the model at the prediction, relating change in one dimension (time) to change in another dimension (*e.g.* overall FRS score). This rate of change defines the shape of an
10 ellipsoid, and thus a distance metric, around the prediction.

Consequently, the set of nearby patient-reported data points may be used to compute a distribution of errors. For each patient-reported data point in this set, the corresponding patient is identified and fit into the model as if they only had as many data points as the patient being predicted (that is, the patient around whose prediction
15 the confidence interval is being evaluated). The disease state at the time corresponding to the patient-reported data point is then predicted from this patient, and the error (difference between the prediction and the actual, uncensored data point) is evaluated. This set of errors produces a distribution, and a desired confidence interval (*e.g.*, a 95% confidence interval) can be computed by evaluating the
20 distribution (*e.g.*, the mean error $\pm 1.96 \sigma$ (standard deviations)). This confidence interval about a particular prediction can be transformed into an overall set of confidence bands around the model for a particular patient by evaluating the error distribution about a set of time-values. The confidence interval can be smoothed by fitting the confidence values to some linear or nonlinear function of time, producing
25 confidence bands that are less sensitive to areas of the model space where there are fewer known data points.

This method is further explained with reference to FIG. 4. In step 402, a predicted value is calculated for a specific patient using the models described herein. In step 404, the n closest data points reported by other patients are then retrieved. For
30 each of these data points (step 406), a data set is obtained in step 408 from the relevant patient (*i.e.* the patient who previously reported the data point selected in step 406). This data set can be limited to the number of data points available for the specific patient. For example, if the specific patient has entered 50 data points, only 50 data points will be retrieved for each of the other patients, regardless of how many

data points may be available. The use of a comparably sized data set permits an accurate estimation of error based on similar conditions. Each data set is used to calculate a predicted value with the model (step 410). In step 412, the error between the predicted value and the actual patient-reported value is calculated. In step 414, a
5 distribution of error is assembled. This distribution of error is then used in step 416 to calculate a confidence interval.

This confidence interval can be represented graphically as depicted by shading
502 in FIG. 5 and may also be used to evaluate the significance of disease
interventions.

10 In addition, the efficacy of a specific treatment can be evaluated based upon examining the disease state (*e.g.*, the ALSFRS-R score) for a set of patients receiving that treatment and comparing the disease state to the expected model state in the absence of such treatment. Efficacy of the treatment may be expressed as a change in the disease state metric responses (*e.g.*, “mean of five point greater ALSFRS-R score
15 than expected at six months”) or as a change in the progression rate (*e.g.*, “median twenty percent less loss of function at four months”).

Pathological Disease Modelling Algorithms

The invention also provides a pathological modeling algorithms for a disease, which relies on an understanding of the underlying pathology of the disease, and can
20 be used to make predictions about a subset of a population or about an individual.

This model describes the state of the disease for some patient at any point in time as a simplified pathological model. For example, in ALS, a finite array representing motor neurons along the spinal cord and degrees of neuron death can serve as the pathological model, as depicted in FIG. 6.

25 For Traumatic Brain Injury (TBI), a connected graph representing regions of the brain and damaged connections (weights on the edges of the graph) or damaged regions (values of the nodes of the graph) can serve as the pathological model.

This model relies on a bidirectional mapping between some user-provided data set and the pathological model. For example, as depicted in FIG. 2, individual
30 questions in the ALSFRS-R or ALSFRS-EX (generically, ALSFRS) can be mapped to individual vertebra regions, and thus a description of the specific damage to the motor neurons can be derived from a set of ALSFRS responses. Similarly, a modeled state of the motor neuron damage may be mapped back to a set of ALSFRS responses.

Similarly, in TBI, individual questions in the Craig Handicap Assessment and Reporting Technique (CHART), the Disability Rating Scale (DRS), the Level of Cognitive Functioning Scale (LCFS), and other published surveys can be used individually or in combination and may be mapped to specific regions of brain trauma or damage.

The mapping can be achieved by establishing an association between individual components (neurons, brain regions, cell types, protein pathways, organs, or other physiological components, depending on the disease) to observable functional or diagnostic effects as communicated by a patient survey, a caregiver survey, laboratory test results as communicated by the patient or a caregiver, or other data provided by the patient or a caregiver. The associations can be established based on existing published or as-yet-unpublished literature, laboratory experimentation, or existing physiological models.

The progression or treatment of the disease can be modeled by establishing factors for the change of the pathological model. For example, in ALS, a set of ALSFRS responses may map to the motor neurons along a certain set of vertebrae being fifty percent functional. Another set of ALSFRS responses may map to those same motor neurons being ten percent functional, and another set being forty percent functional. Based upon a set of these associations, derived from a set of patients used as a training set, data about such population being obtained via a set of surveys on a web site, entry via handheld devices, or other means, a dynamic model can be developed. For example, as, depicted in FIG. 6, in ALS, a Markov model or Hidden Markov model can be trained using the Baum-Welch algorithm to determine an optimal transition matrix describing the decay of neurons in the pathological model. The Baum-Welch algorithm is described in publication such as Leonard E. Baum et al., A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, 41(1) Ann. Math Statist. 164-71 (1970). Additionally or alternatively, a genetic algorithm can be invoked to determine optimal rates of decay of neurons in the pathological model in ALS, or to determine rates of functional adaptation due to retraining of regions of the brain in TBI.

As discussed above, and in the context of FIG. 4, a confidence interval can be determined from the actual error rates of the model. To determine the error bounds around a particular prediction, all known patient outcomes (and thus, associated pathological states) within a certain radius of the prediction can be considered, or

alternatively, the n nearest patient-reported outcomes for some positive integer n . In either case, a valid definition of radius (distance metric) can be derived from the change weights or transition probabilities of the model.

Consequently, the set of nearby patient-reported outcomes can be used to
5 compute a distribution of errors. For each patient-reported data point in this set, the corresponding patient is identified and fit into the model as if they only had as many data points as the patient being predicted (*i.e.*, the patient around whose prediction the confidence interval is being evaluated). The disease state at the time corresponding to the patient-reported data point is then predicted from this censored patient, and the
10 error (difference between the prediction and the actual, uncensored data point) is evaluated. This set of errors produces a distribution, and a desired confidence interval (*e.g.*, a 95% confidence interval) can be computed by evaluating the distribution (*e.g.*, the mean error \pm 1.96 standard deviations). This confidence interval about a particular prediction can be transformed into an overall set of confidence bands
15 around the model for a particular patient by evaluating the error distribution about a set of time-values. The confidence interval can be smoothed by fitting the confidence values to some linear or nonlinear function of time, producing confidence bands that are less sensitive to areas of the model space where there are fewer known data points.

This confidence interval can be represented graphically as depicted in FIG. 5
20 and can also be used to evaluate the significance of disease interventions.

Such a model of disease progression, coupled with such a bidirectional mapping to functional responses or descriptions of the disease, can be used to make predictions as to not only the general course of the disease, but also as to the expected need for (or probability of need for) particular interventions, or the probability of a
25 patient experiencing specific symptoms or outcomes. For example, in ALS, the probability that a patient will need a wheelchair, or a vent, or a feeding tube, can be expressed over time (see FIG. 3 herein). Alternatively, the point in time at which a patient will most likely need an intervention such as one of these, subject to some confidence interval, can be predicted.

30 In addition, the efficacy of a specific treatment can be evaluated based upon inferring the state of the pathological model (based upon observing the functional or diagnostic responses) for a set of patients receiving that treatment and comparing the model state to the expected model state in the absence of such treatment. Efficacy of the treatment can be expressed either as a change in the pathological model (*e.g.*,

“twenty percent reduction in the rate of neuron degradation over a period of four months”) or as a change in the resulting functional, symptomatic, or diagnostic responses (*e.g.*, “thirty percent extension in the median time to requiring the use of a wheelchair”).

5 Data Correction

Another example of this iterative improvement to the model is to correct for the time offset for each patient. Depending on the disease and the specifics of the model, patients may be reporting their date of onset incorrectly due to improper recollection, the imprecise definition of 'first symptom', or the fact that some disease variants may have more obvious first symptoms than others. A correction offset to the onset date for each patient (effectively adding Q to the time value of each data point, where Q is some real-valued number) may reduce the least-mean-squares error of the model. Given that the patient's corrected onset date must be before the date of diagnosis, the optimization for Q may be performed on some bounded interval, *e.g.*,
 15 $Q = \{q \in R : q \geq -d, q \leq d\}$ where d is the time interval between stated onset and diagnosis. This optimization may be performed using a numerical method such as Brent's method. Brent's method is discussed in publications such as R.P. Brent, Algorithms for Minimization without Derivatives (1973). The model function can be re-fit to the new set of parameters in this instance as well, and this process can be
 20 repeated one or more times, until some stopping conditions (*e.g.* convergence, or lack of further change in the value of Q) are met.

Additionally, missing data algorithm (including regression or expectation maximization (EM) algorithms) can be used to generate complex datasets for fitting. These datasets include flags for imputed values to allow sensitivity analyses that
 25 estimate the impact of missing value imputation on model-based forecasts and predictions.

Model Improvement

Exploratory data analysis techniques can be utilized to explore which variables and combinations of variables are associated with the response (outcome, event,
 30 disease state) of interest. A sample, or training dataset is analyzed using simple multiple regression methods (in small, well-behaved samples) or exploratory methods for very large datasets, with or without missing values. In one example, the CHAID (CHi-squared Automatic Interaction Detector) method, involves analyzing and

ordering every possible combination of attributes. The analyst guides the analysis via parameter settings in the CHAID algorithms to identify which attribute is the most important predictor, which attributes have no predictive value, and how attributes combine or interact together to help in prediction of the response.

- 5 In cases where groups of patients are expected to have different response processes (*i.e.*, where different variables are influencing and determining responses) cluster analysis methods may be used to optimize group identification and assignment. Both agglomerative (bottom-up) and divisive (top-down) clustering algorithms, with control over distance measures (which guide the measurement of
- 10 similarity of the cases being considered for group assignment), can be used.

Linear statistical models, such as logit, probit, and proportional hazards can be utilized to produce forecasts and estimated responses.

- Model fit can be improved through the use of Neural Networks, or non-linear statistical models of response. These models are appropriate for pattern recognition
- 15 and modeling when clusters of processes and sub-processes must be taken into account to optimize forecasts and estimates of response.

Prediction of Disease Progression:

- The disease modelling algorithms herein can be used to provide personalized predictions of a particular patient's experience. For example, upon diagnosis with a
- 20 disease, the patient can enter information about the themselves, the disease, and one or more medical conditions. This information is then fed into the algorithms to determine where the patient is along the progression of a disease. For example, as depicted in FIG. 6, the algorithms can predict what an ALS patient's ALSFRS-R value will be at a give point in the future. Likewise, the algorithms can predict when
- 25 certain events will occur, such as confinement to wheel chair or use of a ventilator. Such predictions can be presented to the patient with estimations of confidence in the prediction.

- Referring to FIG. 3, an exemplary user interface 300 is provided for viewing and refining a prediction of disease progression. An icon 302 represents the patient
- 30 controlling the system. The icon 302 includes several color-coded boxes 304, which represent the status of various body systems or regions (*e.g.*, the legs, the spine, and the eyes).

The user interface 300 also includes a population chooser interface 306 for refining the prediction by expanding or contracting the population on which the

prediction is made. For example, a patient can initially view a prediction based on all patients within a community (*e.g.* all patients with ALS). The patient can then alter one or more parameters such as age, gender, race, ethnicity, genotype, etc. The predictions can be updated in real time as the population is altered. In the depicted
5 example, the user can alter the population by sliding one or more sliders 308 to adjust the relative importance a factor such as profile (*e.g.*, age, gender, race, ethnicity, socioeconomic status), genome, disease, function (*e.g.*, as assessed by the ALSFRS-R scale), interventions (*e.g.*, medications consumed), and symptoms (*e.g.*, dysphagia).

The user interface 300 can display icons 310a-310f for one or more patients
10 that are similar to the patient. Icons 310 can be updated as the patient alters the population using population chooser interface 306. The patient can “drill down” to view specific details and profiles of one or more patients, for example, by clicking on one of the icons 310.

The user interface 300 can also include one or more charts 312, 314 depicting
15 predictions of the progression of the patient’s disease. Chart 312 depicts the probability of the patient either (i) recovering from the disease, (ii) living with the disease, or (iii) dying over a twenty-five year period. Chart 314 predicts the probability of the patient requiring assistive devices such as a feeding tube, a wheelchair, or a ventilator over the next twenty-five years.

The user interface can include a graphical element (not shown) that depicts the
20 reliability of the prediction. For example, the graphical element can be modelled after traffic light. A red light can indicate that the prediction lacks a certain level of statistical significance. A yellow light can indicate that the prediction has an intermediate level of statistical significance. A green light can indicate that the
25 prediction has an acceptable level of statistical significance.

The invention can also compute the effect of various stochastic and probabilistic events. For example, the invention can display two different predictions. The first prediction displays the progression of the patient’s disease if the patient develops pneumonia; the second prediction displays the progression of the patient’s
30 disease if the patient does not develop pneumonia. The invention can also display advice on preventing pneumonia.

The invention can also incorporate the probability of such events into the predictive model. This can be accomplished, *e.g.*, through the use of swarm or multiple agent simulation based on known state transition probabilities, as expressed

in Markov chains. Sample measurements can then be taken at arbitrary points in time to determine probabilities of outcomes based on certain criteria. Such criteria can be controllable (*e.g.*, receiving a certain intervention) or uncontrollable (*e.g.*, developing pneumonia).

- 5 Additionally, the invention can simulate the effect of earlier actions that were either taken or not taken. For example, a patient can display the predicted disease progression for colon cancer if the cancer was detected two years earlier. Such a simulation can have a powerful effect on the patient's friends and family.

Verification of Predictions

- 10 In order to further refine the predictions, the methods described herein can include ways of verifying the accuracy of prior predictions. For example, if the algorithm predicts that a patient will be confined to a wheel chair by February 1, 2009, the algorithm can send an email to the patient on or about this date to determine whether this prediction was accurate. Additionally or alternatively, the patient can
15 continue to provide updated data to the algorithm that minimizes the need for follow up emails.

 This newly acquired data is added to the population data and is reflected in further revisions of the predictive models.

Detection of Disease Subgroups

- 20 The invention also enables the detection of rare disease subgroups. For example, certain genotypes exhibit increased or decreased resistance to various diseases. Additional genotypes can be identified by detecting a group of patients that deviate substantially from the predicted disease progression and analyzing the genotypes and other data related to the patients.

- 25 Identification of New Interventions and Off-Label Uses of Medications

- The invention also enables the identification of new interventions and off-label uses of medications. Such interventions and off-label uses can be effected by analyzing data for a population having a disease, identifying patients who experience an improvements in disease progression or symptom severity as a result of an
30 intervention, and identifying the intervention. Given the potentially large size of patient communities, the invention is of particular value to pharmaceutical researchers looking to identify off-label uses of existing medications.

Software/Hardware Implementations

A web-based data-processing system can be used to implement the invention described herein. Web-based data-processing systems are well known in the art and can include a client computer and a server computer. The client and server computers
5 can be coupled to each other over the Internet. Alternatively, the client and server computers can be coupled to each other over an intranet, for example, behind a firewall of a private corporate network. The private corporate network can be the network for a private hospital.

The client computer can include a client software program for executing
10 software applications. The client software program can be an Internet browser such as INTERNET EXPLORER®, available from Microsoft Corporation of Redmond, Washington, FIREFOX®, available from the Mozilla Foundation of Mountain View, California, or OPERA®, available from Opera Software AS of Oslo, Norway. The Internet browser can display content encoded in a variety of standards such as Hyper
15 Text Markup Language (HTML), and FLASH®, AIR®, and ACROBAT® platforms available from Adobe Systems of San Jose, California. User interfaces can include standard web input elements such as text boxes and toggle buttons for entering text and selecting options. The client computer can include input devices, such as a mouse, keyboard, or touch screen for entering information into the user interface.

20 The client computer need not be a personal computer *per se*, but rather encompasses devices such as handheld devices, personal digital assistants, and cellular phones. Mobile devices advantageously allow for more frequent data collection as well as well as reminders for patients to engage in an interventions such as consumption of medication. Suitable mobile device can be specifically constructed
25 for the methods described herein or can be existing mobile devices such a smart phones available under the BLACKBERRY® trademark from Research in Motion Limited of Waterloo, Ontario, the PALM® trademark from Palm, Inc. of Sunnyvale, California, and the IPHONE™ trademark from Apple, Inc. of Cupertino, California.

The user interface can also be a text-based interface. For example, the server
30 can send a text message or an email to a cellular phone or a smart phone asking how the patient is feeling. The patient can respond with an appropriate answer.

Likewise, the user interface can be an audio interface in which the server periodically places a telephone call to the patient asking how the patient is feeling.

The patient can respond verbally, which will be then processed according to known voice recognition software.

The server computer can include a server software program including a web server, for example, Apache Server, and an application server, for example, Cold Fusion Application Server. The server computer can include a database server or engine for encoding and storing data. Suitable database software includes include DB2® and INFORMIX®, both available from IBM Corp. of Armonk, New York; MICROSOFT JET® and MICROSOFT SQL SERVER®, both available from the Microsoft Corp. of Redmond, Washington; MYSQL®, available from the MySQL Ltd. Co. of Stockholm, Sweden; ORACLE® Database, available from Oracle Int'l Corp of Redwood City, California; and SYBASE®, available from Sybase, Inc. of Dublin, California.

The client software program can be used to provide a user interface for entering personalized data related to a patient, for example, a patient diagnosed with ALS. The personalized data can include patient name, sex, and age. The personalized data can include a medical condition metric, for example, whether a patient is feeling great, good, fair, poor, or awful. The personalized data can be submitted to the server software program and the server software program can receive the personalized data.

The server program can store the personalized data in memory on the server computer. The memory can be used to store a data structure including entries for the personalized data. The data structure can be a structured data file or a relational database.

The server software program can analyze the data, for example, using function calls executing on a microprocessor. The server software program can generate a graphical element for representing the personalized data and send the graphical element to the client software program. The graphical element can be sent over the Internet 162 and received by the client software program. The client software program can display the graphical element.

The graphical element can be generated and sent as an image or as a series of values for constructing the graphical element. The image can be sent to the client software program, which can display the image. Alternatively, a series of values can be sent to the client software program, which the client software program can use to construct and display the graphical element. For example, a plug-in executing in an

Internet browser can be used to construct and display the graphical element. The plug-in can include special controls for interacting with the graphical element, including sliders for moving medical condition metrics.

The server software program can also store, analyze, generate, and send to the client software program medical outcome correlations for relating aspects of the medical condition, as further explained herein.

Administrative Tools

The invention includes an administrative tool for use by the scientific and medical staff to evaluate the models used to assess the progression and/or severity of disease. Using statistical techniques, the utility of every predictor variable in the database can be assessed by ranking the predictors by their R^2 values.

Examples

Example 1: Wheelchairs in ALS

A 55-year old male diagnosed with an inherited form of amyotrophic lateral sclerosis enters data stating that he has been tested for a mutation of the super-oxide-dismutase-1 gene (SOD1) known as A4V (Alanine for Valine substitution at point 4). He enters data about his disease progression to-date using a self-report functional outcome scale on his profile in the online community.

An algorithm compares the likelihood of the patient reaching a given clinical milestone (e.g. needing a wheelchair, needing a ventilator, needing to use assistive technology to communicate) by creating a model comparing him to other patients with ALS that are similar in background and also have an A4V mutation.

The algorithm can be based on the ALSFRS. Typical rates of progression are linear with most patients progressing at a rate of 1 point per month. However, analysis of the dataset suggest patients with an A4V SOD1 mutation progress at a rate of 3 points per month.

The patient is presented with the option to see when certain health outcomes are likely to occur. If he decides to see the predicted results, he is shown a description of the system we have used to estimate the outcome. He then clicks on "proceed" and is shown his current progression plot with an overlaid curve progressing from the most recent datapoint / present time, to the predicted health outcomes. Rather than a simple line, the curve presents bands of varying widths according to (i) the quantity

and quality of data provided by the individual and (i) the quantity and quality of data provided by other individuals like him in the system.

As the time of the predicted outcome approaches, the user is sent a private message asking him to validate the accuracy of the prediction made with regards to his health outcome, *e.g.*, “In the past, you used our predictive outcome system to help you understand when you might need a wheelchair. At the time, our model of your disease progress suggested your disease state (as measured by your ALSFRS) would be 22 and you might need a wheelchair around 5 months from now. Please answer the FRS questionnaire. And, do you now use a wheelchair? If so please click “yes” and let us know from what date you started using a wheelchair. If not, please click “no”. We will ask you again in 3 months’ time.”

Positive feedback gained from members decreases the confidence intervals surrounding predictions for a similar group of patients, *i.e.* in this example, future 55-year old male ALS patients with a SOD1 A4V mutation will see a narrower confidence interval around the predicted datapoint of needing a wheelchair. Negative feedback will lead to an increased confidence interval.

Members of scientific staff can evaluate the quality and confidence inherent in a particular model through use of a system tool viewable only by administrators of the website. Models with consistently poor feedback can be examined in detail and altered manually to improve performance.

Example 2: Excessive gambling in Parkinson’s disease

A 75-year old male diagnosed with Parkinson’s disease enters data stating that he has recently been prescribed the drug MIRAPEX® (pramipexol) at a rate of 2 mg per day. He enters data about his treatment regimen using a data-entry module which records his drug regime on his profile in the online community. He enters data that he has had a previous history of alcohol abuse, depression, and gambling.

An algorithm compares the likelihood of the patient reaching a given clinical milestone (*e.g.*, developing a known side effect from the drug (pathological gambling), developing tolerance to the drug and needing a higher dose, finding an improvement in his physical health) by creating a model comparing him to other patients with Parkinson’s disease that are similar in background and are also taking the same drug at a similar dosage.

The patient and other patients in the population enter data about at least two medical condition metrics. First, the patients self-report the severity of their

Parkinson's Diseases by using a scale such as the Unified Parkinson's Disease Rating Scale (UPDRS). Second, the patient enter metrics to track problem gambling, the known side effect of pramipexol. Various scales exist to quantify problem gambling including the South Oaks Gambling Screen (SOGS), the Diagnostic and Statistical
5 Manual of Mental Disorders (DSM-IV), and the Canadian Problem Gambling Severity Index (PGSI).

The patient is presented with the option to see whether certain health outcomes are likely to occur. If he decides to see the predicted results he is shown a description of the system we have used to come to estimate the outcome. He then clicks on
10 "proceed" and is shown the current likelihood of experiencing side effects from the drug on the basis of known data from the clinical literature and/or from other members of the site. Rather than a simple number or percentage chance, the patient is presented with a spectrum of likelihood of varying widths according to (i) the quantity and quality of data provided by the individual and (ii) the quantity and quality of data
15 provided by other individuals like him in the system.

As the time of the predicted outcome approaches, the user is sent a private message asking him to validate the accuracy of the prediction made with regards to his health outcome, *e.g.*, "In the past, you used our predictive outcome system to help you understand whether you might experience a known side effect of MIRAPEX®,
20 excessive gambling. At the time, our model of your disease progress suggested you had around a 20-40% chance of developing excessive gambling in the next 12 months. Have you found this to be true? If so please click "yes" and let us know from what date you started gambling excessively. If not, please click "no". We will ask you again in 3 months' time."

Positive feedback gained from members decreases the confidence intervals surrounding predictions for a similar group of patients, *i.e.* in this example, future 75-year old male Parkinson's disease patients with a history of alcoholism, depression, and gambling will see a narrower confidence interval around the predicted datapoint of developing a gambling problem. Negative feedback will lead to an increased
25 confidence interval.
30

Members of scientific staff can evaluate the quality and confidence inherent in a particular model through use of a system tool viewable only by administrators of the website. Models with consistently poor feedback can be examined in detail and altered manually to improve performance.

Example 3: Rate of progression in Huntington's disease

A 38-year old male with a clinical diagnosis of Huntington's disease enters data stating that genetic testing by his clinician reveals that he has a relatively low number of pathological CAG repeats on the Huntington gene, having only 40 triple
5 repeats. He enters data that he has a high level of education, a high socio-economic status, and a large family able to support him.

An algorithm evaluates the likelihood of him having to be looked after in a nursing home by creating a model comparing him to other patients with Huntington's disease that have a similar number of CAG repeats and also have a high level of
10 education, a high socio-economic status, and a large family able to support him.

The algorithm can use self-reported of functionally ability from other population members as assessed by the Huntington's Disease Rating Scale (HDRS) to predict the patient's future functional ability.

The patient is presented with the option to see when certain health outcomes
15 are likely to occur. If he decides to see the predicted results, he is shown a description of the system we have used to come to estimate the outcome. He then clicks on "proceed" and is shown the current likelihood of having to be cared for in a nursing home on the basis of known data from the clinical literature and/or from other members of the site. Rather than a simple number or percentage chance, the patient is
20 presented with a spectrum of likelihood of varying widths according to (i) the quantity and quality of data provided by the individual and (ii) the quantity and quality of data provided by other individuals like him in the system.

As the time of the predicted outcome approaches, the user is sent a private message asking him to validate the accuracy of the prediction made with regards to
25 his health outcome, *i.e.* "In the past, you used our predictive outcome system to help you understand whether you might need to be looked after in a care home. At the time, our model of your disease progress and family support suggested you had around a 2% chance of needing to be in a care home in the next 12 months. Have you found this to be true? If you did need to be in a care home please click "no" and let us
30 know from what date you started being looked after in a care home. If not, please click "yes". We will ask you again in 12 months' time."

Positive feedback gained from members decreases the confidence intervals surrounding predictions for a similar group of patients, *i.e.* in this example, future 38-year old male Huntington's disease patients a high level of education, a high socio-

economic status, and a large family able to support them will see a narrower confidence interval around the predicted datapoint of needing to be looked after in a care home. Negative feedback will lead to an increased confidence interval.

Members of scientific staff can evaluate the quality and confidence inherent in a particular model through use of a system tool viewable only by administrators of the website. Models with consistently poor feedback can be examined in detail and altered manually to improve performance.

Application to Depression

Some modern theories of depression posit that depression results from cognitive distortions. While all individuals become sad or upset at some points in time, most individuals have the perspective to recognize that such feeling are short-lived. However, individuals with a major depressive disorder are thought by some to lack the ability to recognize recall a time before they entered a depressive episode, and therefore cannot anticipate better times in the future.

The invention described herein are capable of helping persons dealing with depression. Depressed persons can enter their mood or other medical condition metrics into the systems described herein and retrieve graphical representations of these metrics over time. Such a system provides external memory and perspective for the patient.

Furthermore, the inventions described herein can be used by generally healthy individuals in advance of disease. For example, military personnel can record medical condition metrics before deployment to an armed conflict. Such prior medical condition metrics can serve both as a reference point for the military personnel when coping with conditions such a post traumatic stress disorder (PTSD) and to military health personnel seeking to screen for PTSD.

The functions of several elements can, in alternative embodiments, be carried out by fewer elements, or a single element. Similarly, any functional element can perform fewer, or different, operations than those described with respect to the illustrated embodiment. Also, functional elements (*e.g.*, modules, databases, computers, clients, servers, and the like) shown as distinct for purposes of illustration can be incorporated within other functional elements, separated in different hardware or distributed in a particular implementation.

While certain embodiments according to the invention have been described, the invention is not limited to just the described embodiments. Various changes

and/or modifications can be made to any of the described embodiments without departing from the spirit or scope of the invention. Also, various combinations of elements, steps, features, and/or aspects of the described embodiments are possible and contemplated even if such combinations are not expressly identified herein.

5

Incorporation by Reference

All patents, published patent applications, and other references disclosed herein are hereby expressly incorporated by reference in their entireties by reference.

10

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the specific embodiments of the invention described herein. Specifically, although this application periodically discusses the application of the invention to “diseases”, the invention is equally applicable to other medical events such as aging, fertility, and the like. Such equivalents are intended to be encompassed by the following claims.

15

Claims

1. A method for providing real-time personalized medical predictions for an individual patient, the method comprising:
 - providing a database containing patient information for a plurality of other
 - 5 patients including one or more attributes for each patient in the database;
 - constructing a model of a disease based on disease progressions for the plurality of patients;
 - receiving a request from the individual patient, the patient associated with one or more attributes; and
 - 10 making a real-time prediction for the individual patient based on the model and the individual patient's attributes.
2. The method of claim 1, wherein the one or more attributes includes at least one selected from the group consisting of: age, race, ethnicity, gender, height, weight,
- 15 body mass index (BMI), body volume index (BVI), genotype, phenotype, severity of the disease, progression rate of the disease, measures of functional ability, quality of life, interventions, and remedies.
3. The method of claim 1, wherein the disease includes at least one selected from
- 20 the group consisting of: neurological diseases, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson's Disease, Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer blood cancer, fibromyalgia, epilepsy, post traumatic stress disorder, traumatic brain injury, cardiovascular disease, osteoporosis, chronic obstructive pulmonary
- 25 disease, arthritis, allergies, autoimmune diseases, and lupus.
4. The method of claim 1, wherein the data returned includes individual data for one or more members of the set of other patients.
5. The method of claim 1, wherein the data returned includes aggregate data for
- 30 one or more members of the set of other patients.
6. The method of claim 5, further comprising:
 - processing a request from the patient to view individual data.

7. The method of claim 1, wherein the model is based on data for a subset of the plurality of patients, and the method further comprises:

processing a request from the patient to modify a composition of the subset of the plurality of patients.

5

8. The method of claim 7, wherein the composition of the subset of other patients is defined by fuzzy logic.

9. The method of claim 7, wherein modifying the composition of the subset of the plurality of patients includes modifying the range of attributes of patients within the subset.

10. The method of claim 7, wherein modifying the composition of the subset of the plurality of the patients includes modifying the importance of attributes of patients in composing the subset.

15

11. The method of claim 1, further comprising:

conducting a multivariate pattern matching search of data related to the plurality of patients.

20

12. The method of claim 1, further comprising:

calculating a confidence interval for the prediction.

13. The method of claim 12, wherein calculating a confidence interval for the prediction constitutes:

25

selecting a set of reported data points from the plurality of other patients;

for each of the reported data points in the set:

30

obtaining a data set for the corresponding other patient to the reported data point;

calculating a predicted value with the data set and the model;

and

calculating an error between the predicted value and the reported data point;

producing a distribution of the errors; and
calculating a confidence interval from the distribution.

14. The method of claim 13, wherein the set of reported data points includes n closest reported data points to the prediction.

5 15. The method of claim 13, wherein the set of reported data points includes reported data points within an ellipsoid defined by a distance metric.

16. The method of claim 13, wherein a size of the data set for the corresponding other patient is comparable to a quantity of attributes associated with the individual patient.

10

17. The method of claim 1, further comprising:
analyzing an effect of an intervention by measuring a difference between a prediction absent the intervention and a reported outcome with the intervention.

15 18. The method of claim 17, wherein the difference is measured for a plurality of individual patients.

19. The method of claim 17, wherein the difference is compared to the distribution of error.

20

20. The method of claim 17, wherein the difference is compared to the confidence interval for the model.

21. The method of claim 20, further comprising:

25 identifying one or more of the differences that exceed the confidence interval for the model.

22. The method of claim 18, further comprising:

assembling a distribution of the differences for the plurality of individual patients; and

30

computing a standard error for the distribution.

23. The method of claim 12, wherein the confidence interval is calculated with a chi-square test.

24. The method of claim 12, wherein the confidence interval is calculated from a
5 measure of variance of the individual patient's attributes.

25. The method of claim 12, wherein the confidence interval is calculated by
comparing the individual patient's attributes to a model fit for the individual patient
using the model.
10

26. A computer-readable medium whose contents cause a computer to perform a
method for providing real-time personalized medical predictions for an individual
patient, the method comprising:
providing a database containing patient information for a plurality of other
15 patients including one or more attributes for each patient in the database;
constructing a model of a disease based on disease progressions for the
plurality of patients;
receiving a request from the individual patient, the patient associated with one
or more attributes; and
20 making a real-time prediction for the individual patient based on the model and
the individual patient's attributes.

27. A method for providing real-time personalized medical predictions, the
method comprising:
25 gathering patient-submitted information from a community of patients having
a disease, the information including medical condition metrics and intervention data;
utilizing the patient-submitted information to form a model of the disease; and
predicting the progression of the disease in a particular patient by applying
information submitted by an individual patient to the model.

30 28. The method of claim 27, wherein the information submitted by the individual
patient includes a date of onset of the disease.

29. The method of claim 27, wherein the step of predicting the progression of the disease is performed in real time.
30. The method of claim 27, wherein the step of predicting the progression of the
5 disease including providing a confidence interval.
31. The method of claim 27, wherein the step of predicting the progression of the disease including providing a graphical prediction.
- 10 32. The method of claim 27, wherein the graphical prediction is a line chart depicting development of the disease with regard to a rating scale.
33. The method of claim 27, wherein the disease includes at least one selected from the group consisting of: neurological diseases, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Parkinson's Disease, Human Immunodeficiency
15 Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), depression, mood disorders, cancer blood cancer, fibromyalgia, epilepsy, post traumatic stress disorder, traumatic brain injury, cardiovascular disease, osteoporosis, chronic obstructive pulmonary disease, arthritis, allergies, autoimmune diseases, and lupus.
- 20 34. The method of claim 27, wherein the model is a model of the disease's pathology.

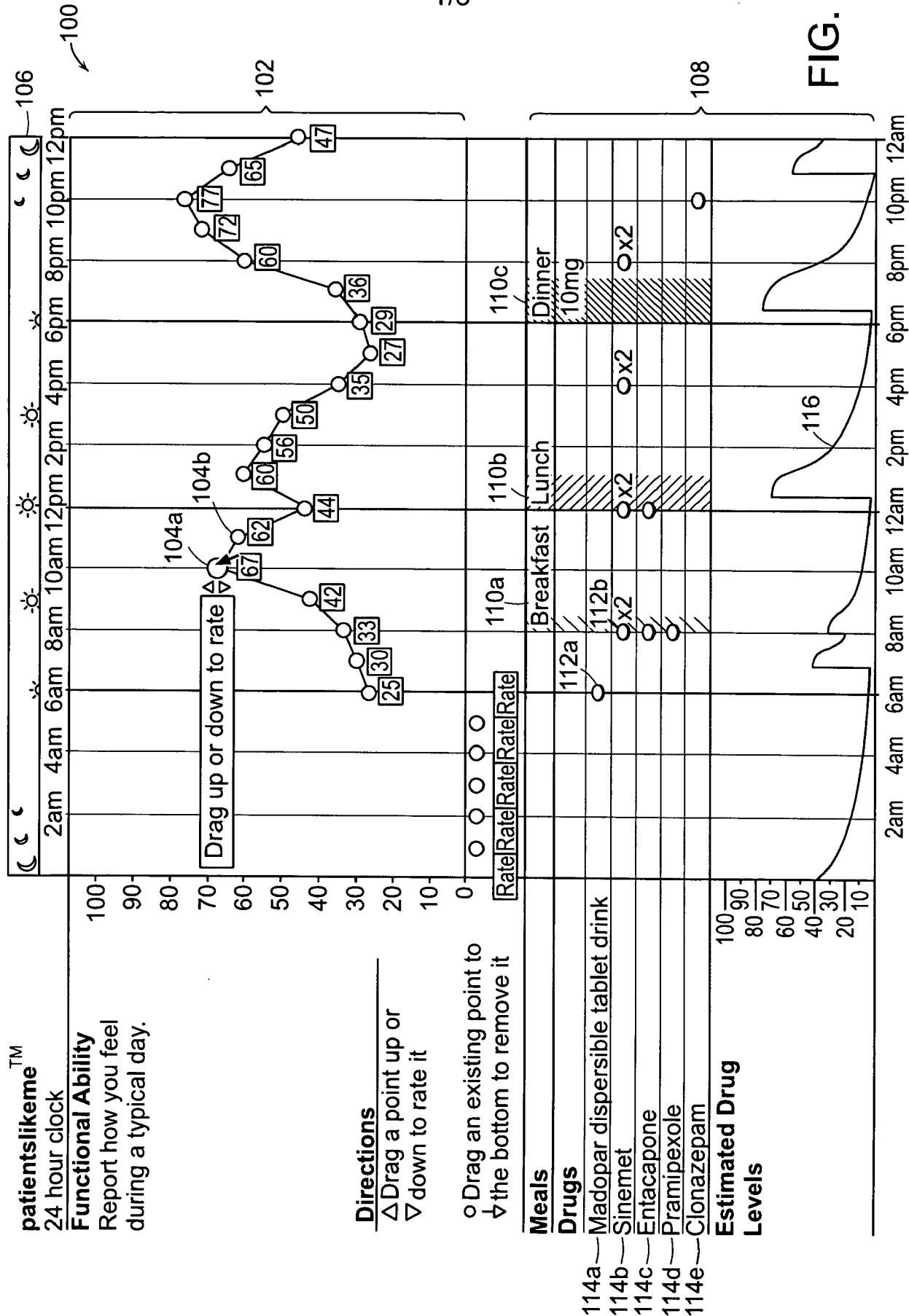


FIG. 1

FIG. 2A	FIG. 2B
------------	------------

FIG. 2

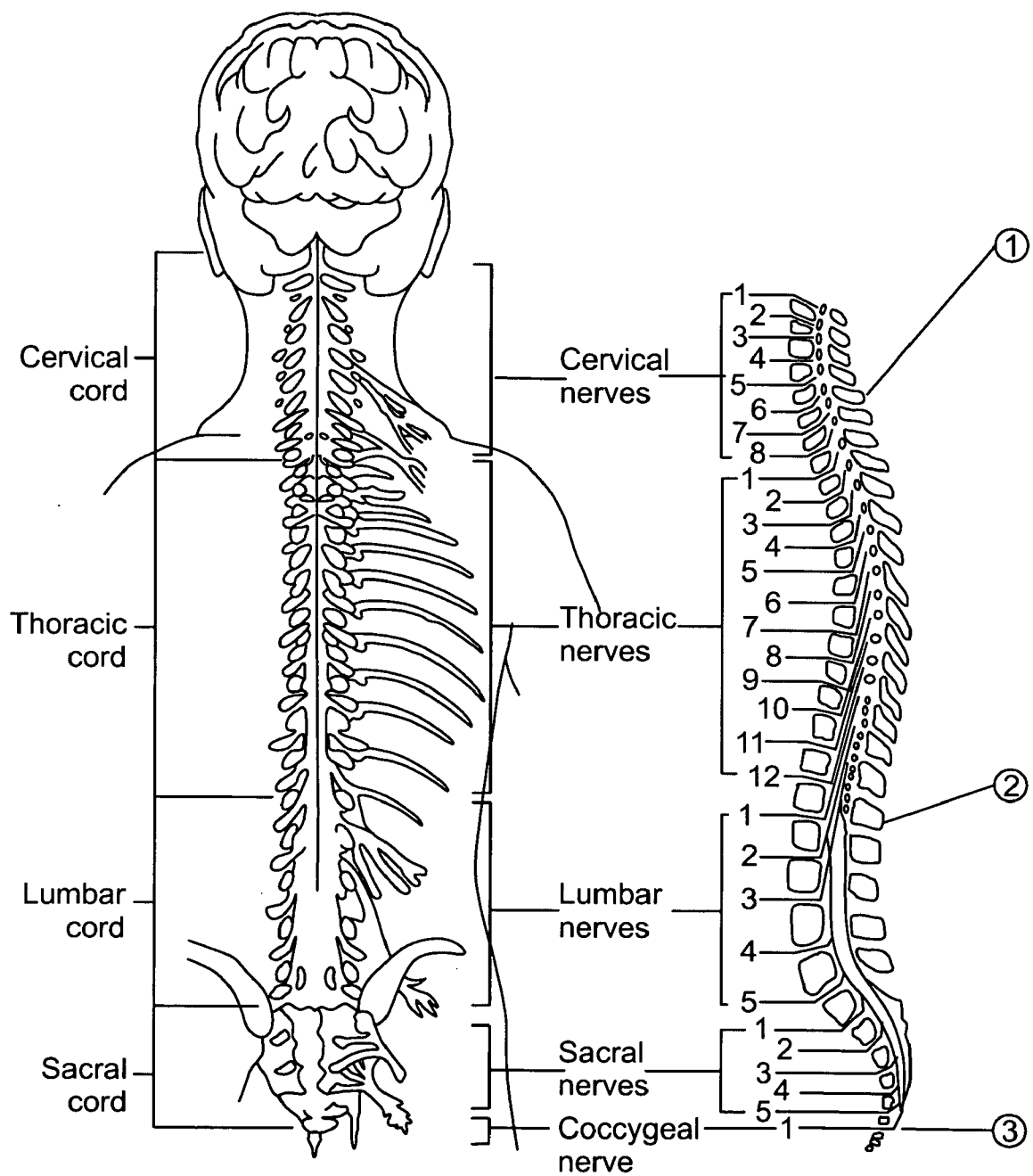


FIG. 2A

3/8

ALS FRS Questions
Mapped to Spine Region

	Spinal Region	Speech	Salivation	Swallowing	Handwriting	Food	Dressing	Turning	Walking	Stairs	Syspna	Orthopena	Respiratory
	C1												
	C2	•		•									
	C3	•	•	•							•	•	•
①	C4		•	•							•	•	•
	C5					•	•	•			•	•	•
	C6				•	•	•	•					
	C7				•	•	•	•					
	C8				•	•	•						
	T1				•	•	•				•	•	
	T2						•				•	•	
	T3						•				•	•	
	T4						•				•	•	
	T5						•				•	•	
	T6						•			•	•	•	
	T7						•			•			
	T8						•	•		•			
	T9						•	•		•			
	T10						•	•		•			
②	T11						•	•		•			
	T12						•	•		•			
	L1						•	•	•	•			
	L2						•	•	•	•			
	L3						•	•	•	•			
	L4						•	•	•	•			
	L5						•	•	•	•			
	S1						•		•	•			
	S2						•		•	•			
	S3												
	S4												
③	S5												

FIG. 2B

FIG. 3A
FIG. 3B

FIG. 3

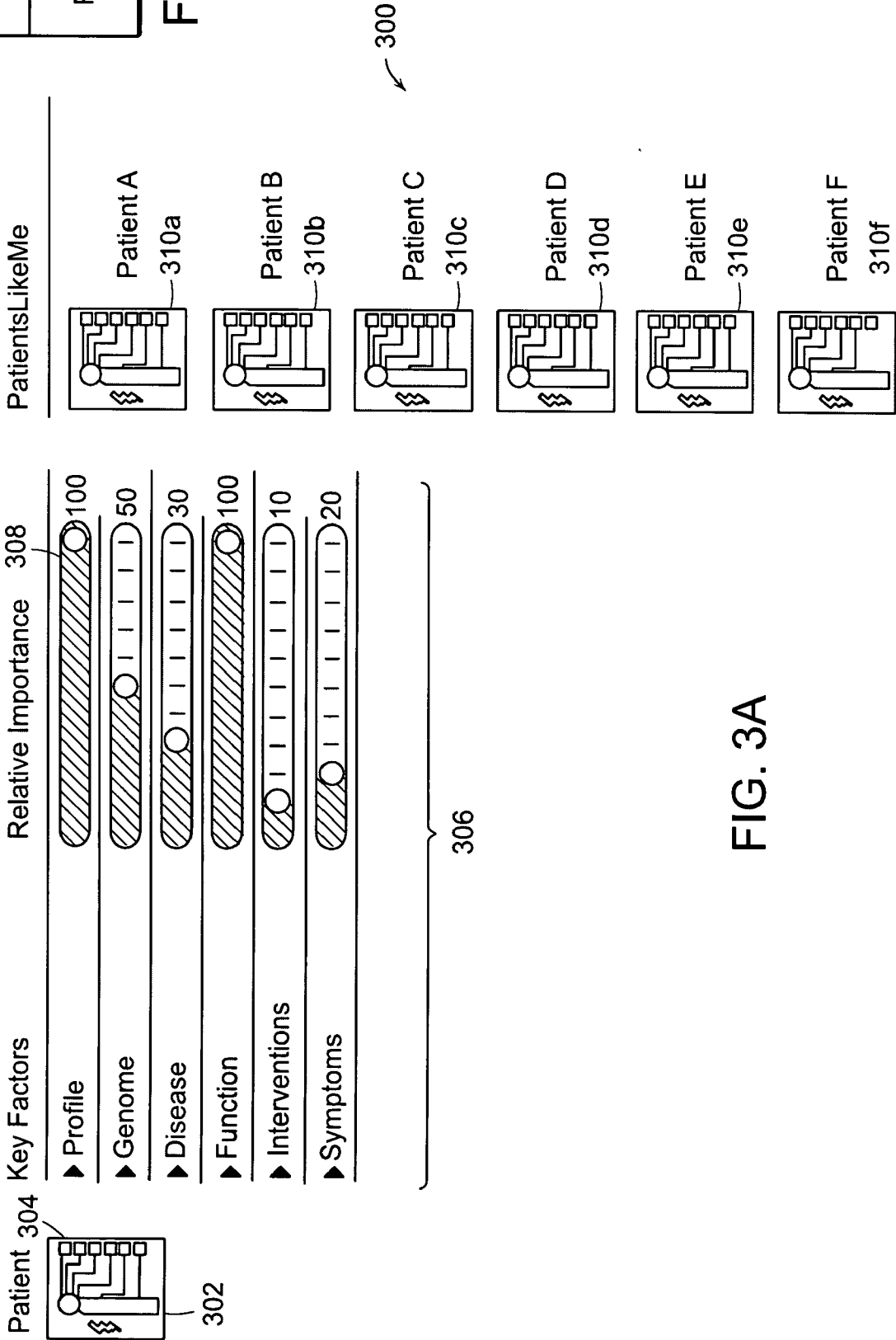


FIG. 3A

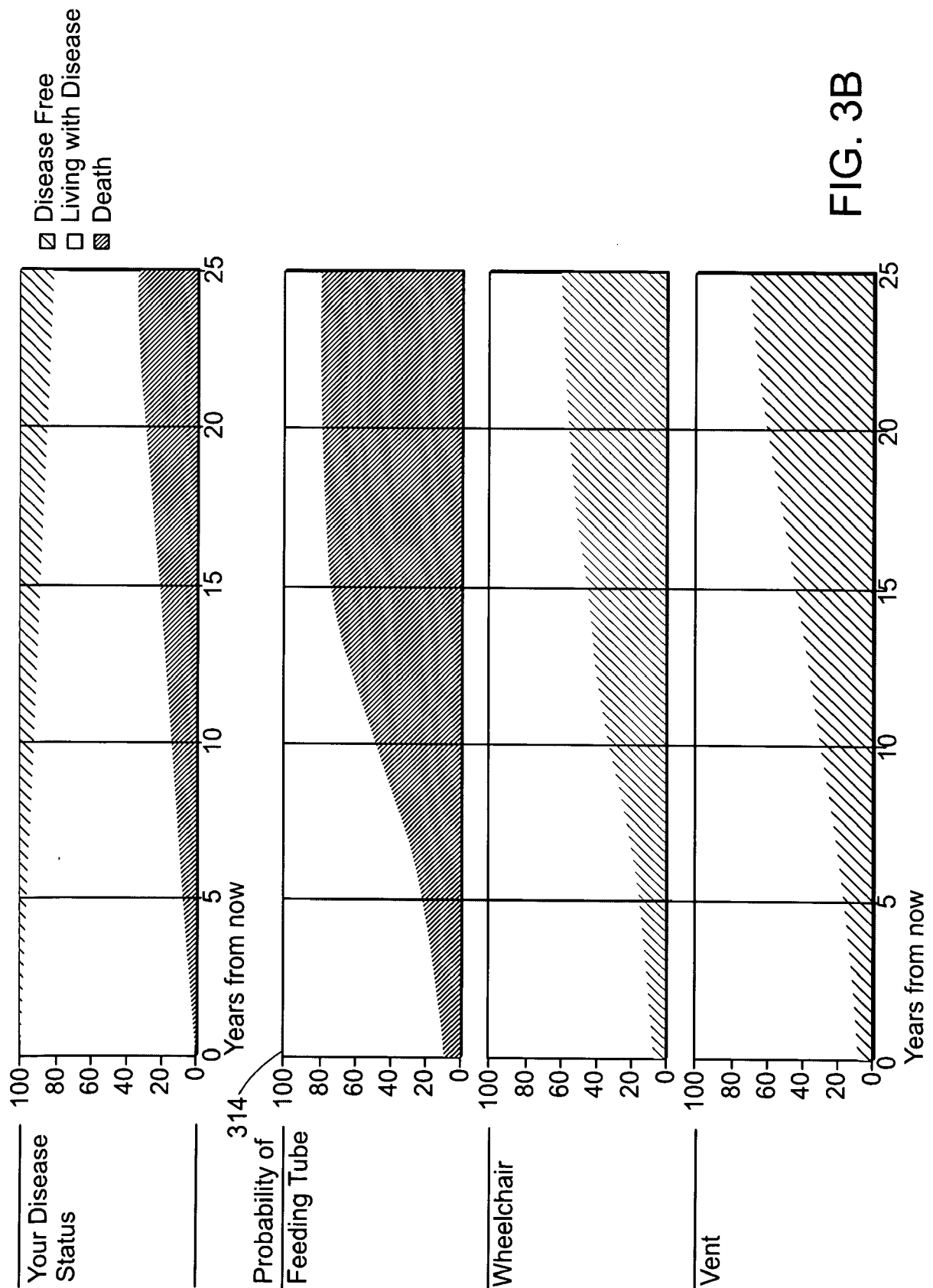


FIG. 3B

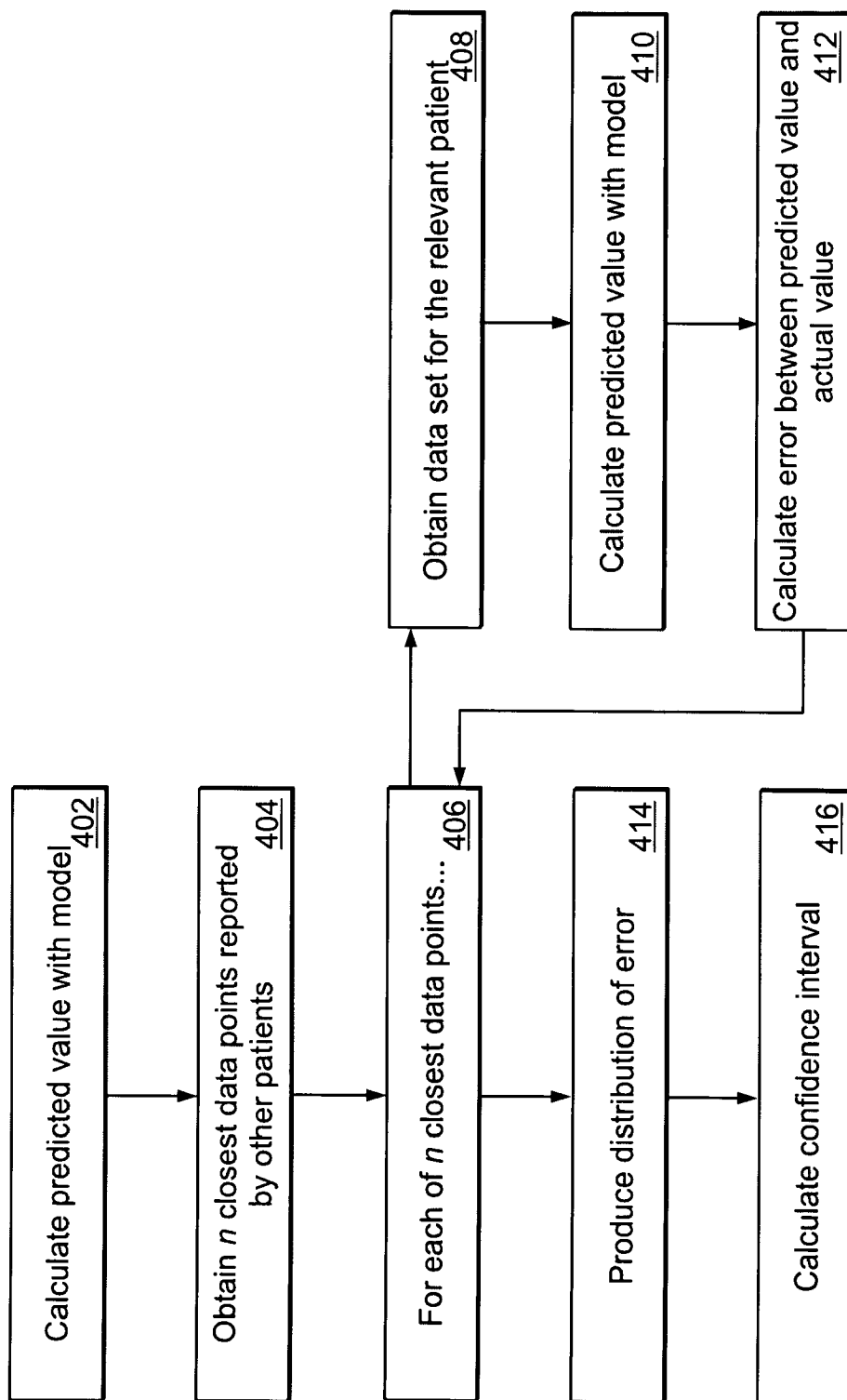


FIG. 4

7/8

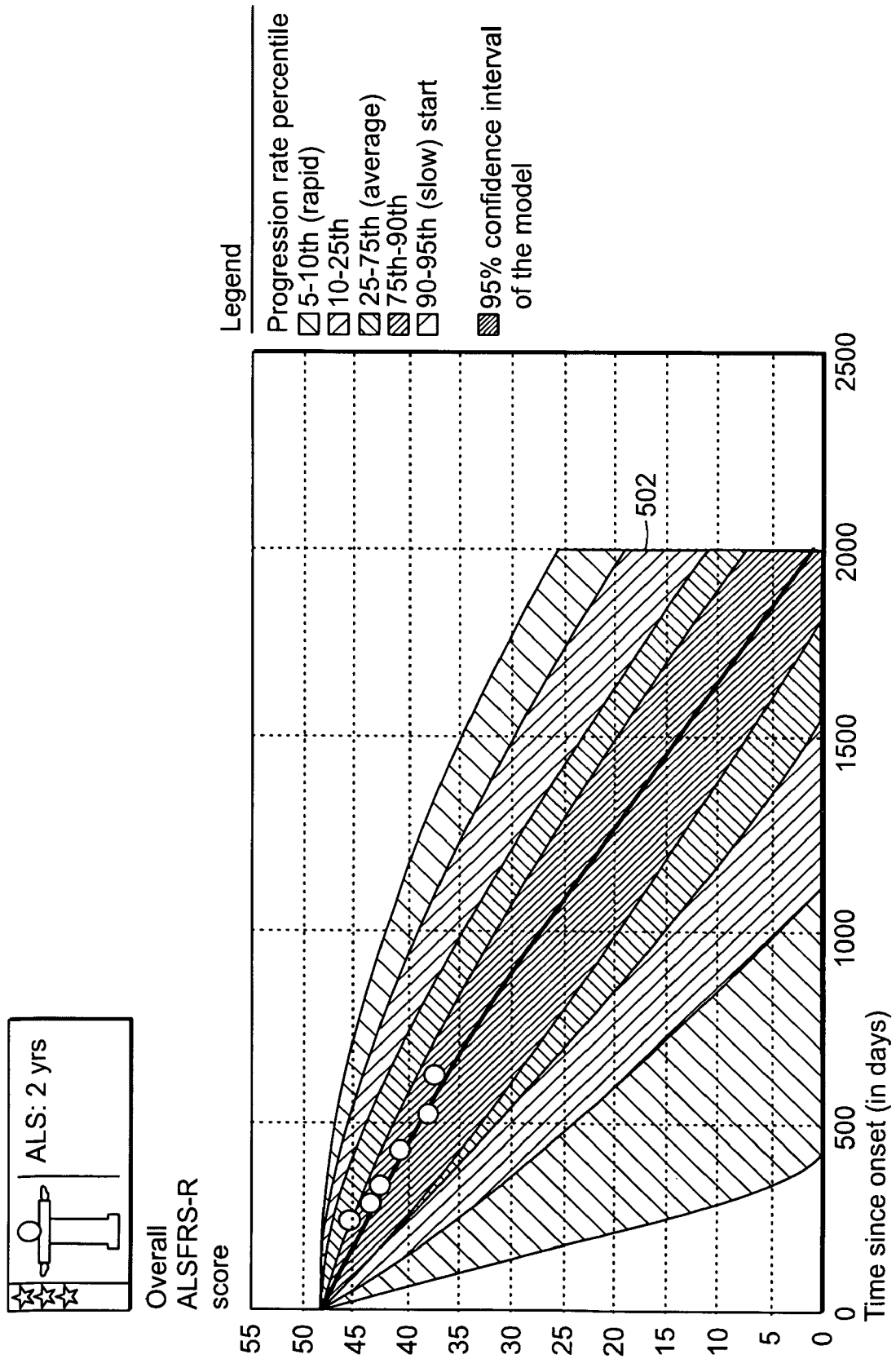
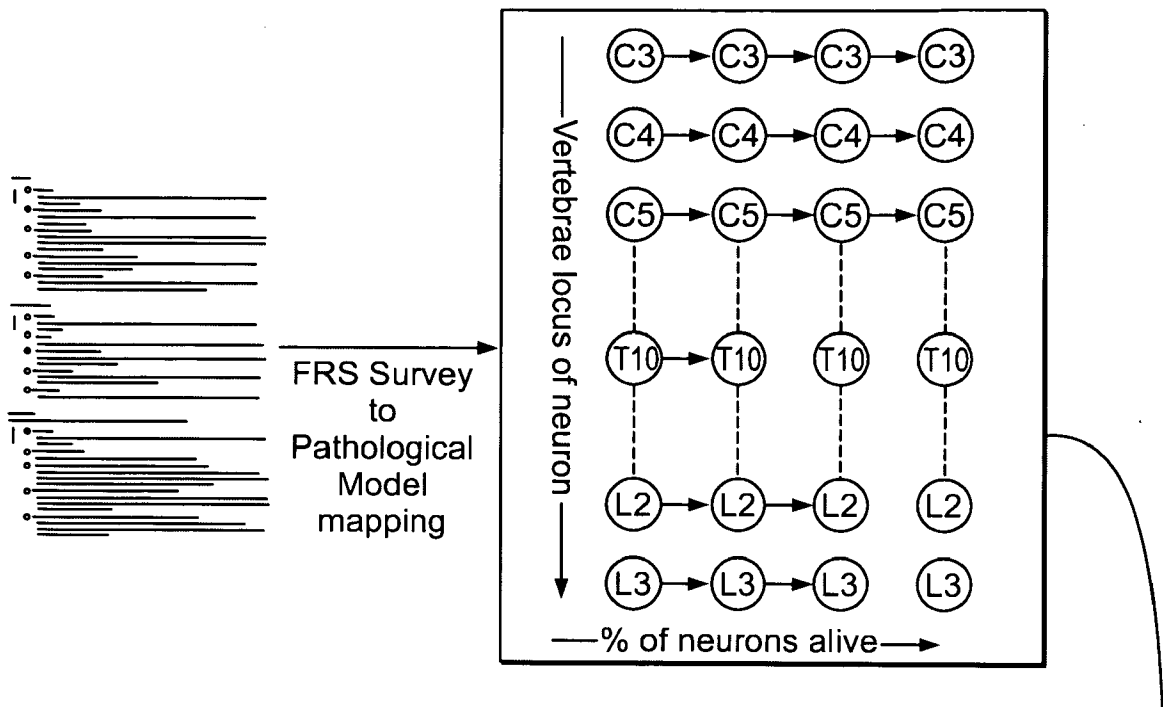


FIG. 5

8/8



Progression of neuron death over course of four months for this patient according to Hidden Markov Model (note further neuron death in L2, L3, T10, C5)

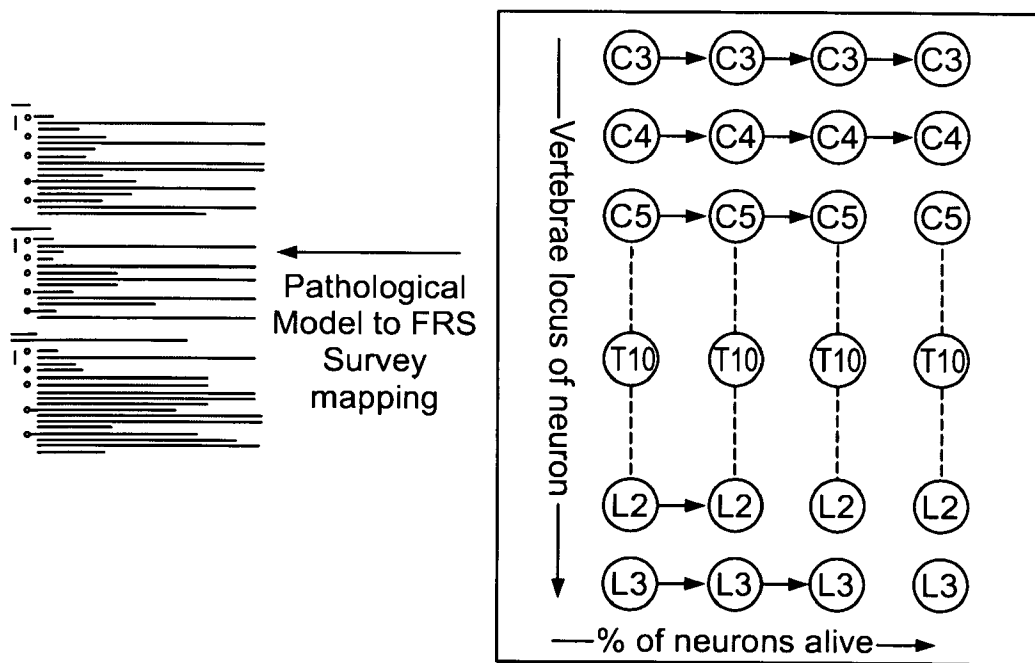


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 08/79674

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06Q 10/00 (2008.04)

USPC - 705/2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

USPC - 705/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 705/1, 3, 7, 8, 500; 700/90, 91, 93

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Electronic Databases Searched: PubWEST (PGPB, USPT, EPAB, JPAB); Google Scholar

Search Terms Used: medical, health, prediction, forecast, tracking, patient, ill, sick, hospital, clients, database, illness, disease, ailment, affliction, HIV, AIDS, cancer, depression, models, representations, simulations, progressions etc.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	US 2004/0015337 A1 (Thomas et al.) 22 January 2004 (22.01.2004), para [0006]-[0009], [0013], [0015]-[0016], [0020], [0031], [0033]-[0035], [0037]-[0038], [0040], [0047]-[0060], [0062], [0066]-[0067], [0076], [0082], [0099], [0103], [0115]-[0117], [0130], [0135], [0137], [0145]-[0146], [0149]-[0150], [0153], [0156], [0159]-[0160], [0164], [0169]-[0170], [0175] and Fig. 5	1-3, 5, 7, 9-11 and 26-34 ----- 4, 6, 8 and 12-25
Y	US 2006/0052945 A1 (Rabinowitz et al.) 09 March 2006 (09.02.2006), para [0033], [0036], [0040], [0072], [0207]-[0208], [0232]-[0233], [0251]-[0254], [0270] and Fig. 11	12-16 and 23-25
Y	US 2005/0080462 A1 (Jenkins et al.) 14 April 2005 (14.04.2005), para [0017]-[0018], [0022], [0032], [0049]-[0050], [0078], [0095]-[0096], [0153], Fig. 10 and 11	17-22
Y	US 2006/0036619 A1 (Fuerst et al.) 16 February 2006 (16.02.2006), para [0109]	4 and 6
Y	US 2004/0122707 A1 (Sabol et al.) 24 June 2004 (24.06.2004), para [0416]	8

☐ Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

14 November 2008 (14.11.2008)

Date of mailing of the international search report

01 DEC 2008

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774