

The Limitations of Big Data in Healthcare

Abstract

It has become commonplace in modern society to collect and process large bodies of data in order to establish trends in the data and thereby identify business opportunities, areas of cost-saving, etc. This article questions whether, in the healthcare context, the data sets from biomarker-type tests are reliable indicators of dysfunction and whether the processing of such data sets can be expected to identify significant areas for improvement in the clinical context which would lead to significant improvements and cost-savings regarding the diagnosis and treatment of morbidity. In particular, there is not yet an accepted understanding of the mechanisms which are responsible for maintaining the body's highly regulated function therefore the idea that the current plethora of diagnostic tests can be accurate measures of the rate of pathological onset of a particular medical condition(s) is only an assumption and may have significant inherent limitations.

The development of the first mathematical model of the autonomic nervous system by Grakov leads us to question basic assumptions upon which modern medicine is based and whether large-scale data mining of available data sets will lead to a better understanding e.g. (i) Is the biomarker the cause or the consequence of the condition? (ii) Does the marker relate to the genotype, phenotype or neither? (iii) How does the brain regulate the autonomic nervous system and the coherent function of the organ networks/physiological systems?

Keywords: Physiological systems; Organ Networks; Biomarker; Non-Genetic; Genomics; Transcriptomics; Proteomics; Metabolomics; Nervous system; Cognitive psychology; Migraine; Raynaud's Phenomenon; Homeostasis

Review Article

Volume 5 Issue 2 - 2017

Graham Wilfred Ewing*

Mimex Montague Healthcare Limited, UK

***Corresponding author:** Graham Wilfred Ewing, Mimex Montague Healthcare Limited, Mulberry House, 6 Vine Farm Close, Cotgrave, Nottingham NG12 3TU, Email: graham.ewing@mmhcl.co.uk

Received: January 31, 2017 | **Published:** March 03, 2017

Introduction

Modern medicine uses a multi-level approach to determine the health of the patient.

- i. The human body is an intensely regulated entity which medicine characterises by a range of physiologically significant parameters often known as 'vital signs'.
- ii. The doctor's examination is based upon a rudimentary assessment of the stability or instability of the patient's systemic stability i.e. of blood glucose, blood pressure, pH (of urine), posture, digestion, elimination, sleep, etc.
- iii. Histopathology tests are based upon the assumption that the body's function can be explained solely by its biochemistry and hence that by measuring levels of suitable biomarkers, that it is possible to use such knowledge to identify problematic health issues in the patient and, ultimately to understand the regulatory mechanisms which control the body's physiology.

It is increasingly recognised that most medical conditions are characterised by a spectrum of pathological changes. A biomarker is often only one of the many pathological changes which accompany the development of a medical condition. In addition, most medical conditions have multi-systemic origins. If so, what is the nature and function of these systems? All medical conditions are multi-pathological and polygenomic. Moreover, genetic changes often occur alongside pathological onset i.e. the genetic changes are the consequence of pathological onset.

In recent papers the author has elucidated upon the mechanisms which the brain uses to regulate the autonomic nervous system and physiological systems i.e. the systems upon which medicine is based. Such knowledge is not just a hypothesis but is supported by a working, commercialised technology known by the brand name STRANNIK. New technologies bring new levels of understanding and so is the case with Strannik. It enables us to make constructive efforts and comments upon efforts by medical researchers to establish better ways to diagnose and treat the many and various medical conditions influencing patient health.

The Limitations of Big Data: The Accuracy of The Data

The problem which is faced by researchers as they seek to turn BIG DATA into smart data is that it starts with the fundamental and erroneous assumption that this huge amount of data can be adapted and lead to real time assistance in disease prevention, prognosis, diagnosis and therapeutics. Consider for a moment whether the data is compatible. Of course there will be progress made which can be presented as case studies using immense computational resources and tools which take data from -omics sources, in particular the levels of particular metabolites, however there appears to be a general assumption that levels of biochemical markers and/or metabolites are accurate determinants although each pathological process and/or medical condition comprises a spectrum of pathological coordinates, each of which is characterised by its genotype and the non-genetic factors arising from lifestyle and/or environmental factors i.e. phenotype.

- a. Is the total level of a particular protein an accurate determinant when considering that the body produces, and may also store, more of a protein than is required or is consumed? We know that proteins may be coiled or uncoiled and that this is a significant although unexplained factor in many pathological processes. Moreover if the protein is uncoiled it will be in an unreactive form i.e. it is only the level of reactive protein which is significant. In many cases the surplus levels of proteins are eliminated by chaperonin enzymes, by natural process of decay, etc. In other cases the excess of a protein may have pathological significance as the protein reacts with other secondary substrates.
- b. There are NO tests which are precise and accurate. There cannot be because every medical condition comprises a spectrum of pathological coordinates. In some patients one pathological coordinate will have greater significance than in other patients and the test will be more or less accurate. In other conditions e.g. Raynaud's phenomenon, there will be a spectrum of pathological correlates. Most medical tests are only relatively accurate under a set of conditions e.g. the HbA1c test can be influenced by light, pH, differing levels of haemoglobin, etc.
- c. Levels of insulin have no effect upon pathological onset. If insulin is the fundamental cause (i) how can diabetes occur in patients with normally functioning pancreases as happens in patients with hysterectomy? [1] (ii) Why doesn't the onset of diabetic complications decline following the therapeutic application of insulin? The onset of diabetic comorbidities e.g. to the heart, kidneys, periphery, etc; continues to develop. This indicates the presence of a more sophisticated regulatory mechanism than can be justified by the prevailing biological model.
- d. The mere act of presenting and/or inserting a needle/sampling syringe causes the innate immune system to respond to the needle and therefore any sample taken is not and cannot be an accurate determinant of the patient's medical condition although in most cases the deviation is not considered to be significant.
- e. Most medical tests consider only a portion of a normal Gaussian distribution curve when comparing test results with numbers having the condition, typically between 10% and 90% within the range 4-8mmol/litre. If between 10% and 90% the results are considered to be confirmation of a positive or negative result; if outside these limits the test results are considered to have questionable validity and indicative of the pathological process e.g. the results of blood glucose levels which confirm whether the patient's reported conditions are normal; therefore by definition every test incorporates a set of assumptions and factors which influence test outcomes to some degree. The problem is that setting such limits incorporates bias and/or errors to the reported results.
- f. The process involves labelling samples, storing samples and loading samples into sampling carousels. At the end of this process the reported results are sent to the patient's GP. The process is vulnerable to errors.
- g. Both the genetic and phenotypic processes which influence pathological onset or progression have to be quantified [2,3]. This is immensely significant when diagnosing diabetes. Erroneous diagnosis could for example lead to the patient being administered insulin when instead the most appropriate intervention would be through a suitably tailored programme involving diet and exercise.
- h. This could be done by monitoring insulin levels however insulin levels rise and fall in a cyclic manner every 3-15 minutes [4], according to demand or need, and according to the level of diabetes; and by measuring the rate of glycation which also rises and falls according to diet. The current method of measuring glycation uses Glycated Haemoglobin, in particular HbA1c, however this can be influenced by variable levels of haemoglobin which often arises in patients with diseases which are influenced by liver dysfunction. It requires patients to be tested for abnormal levels of haemoglobin before being considered suitable (or otherwise) to be given the HbA1c test [5]. Recent research indicates 40% irreproducibility after one month and other shortcomings [6,7]. The test is known to be influenced by pH, levels of sunlight, levels of pO₂, etc. Perhaps there is need for a better test which employs a direct marker rather than the indirect marker HbA1c. Glycated albumin has been proposed as an alternative marker.

Accordingly any system which relies upon the accuracy and validity of the current plethora of biomedical data will inevitably encounter significant problems due to the basic assumptions upon which biomedicine is based and hence the inherent limitations of the technique(s).

Limitations of Big Data: Where and How It Needs to be Applied

If the huge amounts of BIG DATA are to be assimilated to give rational conclusions there has to be a body of data which is compatible and which can be mathematically processed. There has to be a common denominator and data which is compatible with the common denominator. For example it is difficult to compare the numbers of apples with numbers of oranges and should consider comparing the numbers of fruit and so too must be the case as we compare the output from the huge range of biochemical tests. There is a need to gather data in a form which is capable of yielding significant data. Just gathering more and more data of spurious quality is unlikely to yield significant outcomes [8] e.g.

- a. It is difficult to compare blood pressure with a particular biochemical marker because blood pressure is a neurally regulated network of organs more commonly referred to as a physiological system. Emergent pathologies in any of the organs in this physiological system will influence the brain's ability to maintain normal levels of blood pressure.
- b. It is difficult to compare blood glucose with a particular biochemical marker because blood glucose is a neurally regulated network of organs in which blood glucose levels are regulated by the brain i.e. (i) alterations of brain function and consequently of the brain's ability to regulate blood glucose

levels OR (ii) the emergence of pathologies in the network of organs which regulate blood glucose levels, influence the prevailing levels of blood glucose. Moreover it is critically important to consider what is being measured. Are we measuring the expression of insulin (and/or its precursor(s)) more commonly referred to as genotype or are we measuring the rate at which the expressed protein reacts with its reactive substrate Insulin Receptor Protein i.e. phenotype?

- c. Blood Pressure and Blood Glucose are two of the body's physiological systems; the others being blood cell content, blood volume, breathing, pH, osmotic pressure, sleep, digestion, excretion of fluids, body temperature, posture/musculoskeletal system and sexual function. In medicine instability in each of these systems are characterised by the use of the prefixes 'hyper' or 'hypo' e.g. hyperglycaemia, hypoglycaemia, hypertension, hypotension, etc. Moreover instability in any of these systems will inevitably influence stability of other adjacent systems as the brain seeks to maintain best-fit optimum stability (homeostasis).

In order to be consistent with scientific theory it is necessary to consider and/or recognise that there are genetic and non-genetic factors (phenotype) which characterise the rate of onset and progression of every medical condition; that the brain regulates the autonomic nervous system and physiological systems; that pathological onset influences brain function; and that stress in its many and various manifestations influences the brain's ability to maintain optimum or best-fit stability.

The autonomic nervous system comprises the sympathetic nervous system i.e. the physiological response to stress which differs between people according to their many and various lifetime of experiences; and the parasympathetic nervous system which is able to return the body to a stable 'base' level; however as yet there are no reliable or scientifically accepted or valid ways of measuring such parameters.

Diabetes is a wonderfully interesting subject because it throws up a number of anomalies e.g.

- d. It is generally assumed that diabetes is a problem of the pancreas yet there is a phenomena 'non-pancreatic diabetes' [1] which indicates that there are problems with the simplistic assumption that abnormal levels of blood glucose – as a result of pancreatic dysfunction – is the fundamental cause; because diabetes can occur in patients (who have had a hysterectomy) who have normal pancreatic function i.e. the levels of blood glucose appear to be the consequence of systemic dysfunction.
- e. The pancreas is an organ in the network of organs (physiological system) which regulates blood glucose [9-12].
- f. Proteins such as insulin may be coiled and reactive or uncoiled and unreactive. Accordingly it is necessary to understand the factors which influence protein morphology.
- g. In most patients diabetes presents as a combination of genotype and phenotype influencing the function of the cells of the islets of Langerhans [2,3]. This occurs because the genetic expression of proteins (in this case insulin)

occurs BEFORE the subsequent reaction of insulin with its receptor protein i.e. phenotype (genotype and phenotype are comorbidities);

- h. The administration of Insulin does not prevent the development of diabetic comorbidities e.g. cardiovascular disease(s), stroke, renal complications, diabetic retinopathy, etc.
- i. The brain regulates the autonomic nervous system and physiological systems (blood glucose is one of these physiological systems/organ networks) therefore there is a need to understand and model the structural nature of the autonomic nervous system and physiological systems.
- j. In order to characterise any chemical reaction it is necessary to measure the rate at which the reactive substrates react, under set conditions, to form a product. This applies to biochemistry. The reaction conditions, which are often overlooked in medical research, must also be taken into account i.e. the prevailing pH influences pathological onset and/or progression. If pH decreases the levels of essential minerals (magnesium, calcium, zinc, chromium), which are available to support physiological stability, will decline and levels of transition metals (iron, aluminium, lead, mercury, etc) will increase. This sets the preconditions for the free radical processes which result in glycation and which leads to diabetes, increased blood viscosity, elevated blood pressure, the formation of complex lipids, atherosclerosis, etc.

In addition, proteins are highly polar entities with an -NH₂ group and -COOH group. Accordingly, increased levels of acidity must influence the 3-D structure or morphology of proteins and their reactivity.

Yet proteins provide the answer to the problems outlined at the beginning of this text. Proteins are visually active. They release biophotons in the course of their reaction which influence our colour perception which explains why the blood is bioluminescent [13,14] therefore, at least in principle, it is possible to measure changes of brain function and sense/colour perception and use this as a measure of the rate at which proteins (i) are expressed (genotype) and (ii) react with their reactive substrates (phenotype).

What is the Intended End-Product of Researching or Evaluating 'Big Data'?

The idea of BIG DATA was conceived with the intention of linking -omics data comprising genomics, transcriptomics, proteomics, metabolomics, etc; yet it makes fundamental omissions e.g. that protein expression can be influenced by environmental factors, proteins may be coiled and reactive or uncoiled and unreactive, the brain regulates the autonomic nervous system, that inorganic processes influence biological/pathological processes, the data being gathered and processed is of suitable quality and reliability, that the total quantity of a particular protein or other biologically active substrate is significant, and that the fundamental mechanism which is responsible for most lifestyle-related medical conditions is non-genetic i.e. genetic and non-genetic processes are co-morbidities.

The brilliant Russian researcher Dr Igor Gennadyevich Grakov has used cognitive data to create a comprehensive and sophisticated mathematical model of the autonomic nervous system thereby linking cognitive psychology to the reaction kinetics of pathological processes. It includes an understanding of the structural relationship between brain function and the physiological systems, the organs in each of these systems, the cells in each of these organs, and molecular biology i.e. the emergent pathological processes which influence cellular biology and which often influence brain function.

Changes of colour perception have pathological origins. Proteins are visually active. They absorb and emit light during the course of their reaction which is responsible for bioluminescence of the blood. It is this emission of light, in particular, which influences colour perception. Accordingly changes of sense perception, and in particular of colour perception, can be used as an accurate determinant of pathological onset and progression [14].

Such a technology can be applied with diagnostic and therapeutic effect. It uses knowledge of the natural processes by which the brain functions and regulates the autonomic nervous system. It simulates brain function and effectively considers the brain as a black box computer which can be fed sensory input and generates pathological output. This can be applied with diagnostic or therapeutic effect i.e. to determine the pathological conditions influencing the health of the patient, to understand what is likely to happen if they continue their current lifestyle, and to modulate brain function with therapeutic effect.

Conclusion

Initial research indicates that SVS is circa 2-23% more accurate than the range of diagnostic technologies against which it was compared and that SLT is circa 83-96% effective depending of course upon the nature and extent of the conditions to be treated.

Recent published papers by the author illustrate how SVS can be used to determine the complex range of pathological coordinates in patients with diabetes, diabetic comorbidities, cardiovascular disease [15], migraine [16], Raynaud's phenomenon [17], Alzheimer's disease [18], etc; and how SLT can be used with therapeutic effect in cases of diabetic circulatory issues (ulcers), high blood pressure, abnormally low cytokines levels [19], migraine [20], headache, depression, sleeping disorders [21], etc.

References

- Appiah D, Winters SJ, Hornung CA (2014) Bilateral Oophorectomy and the Risk of Incident Diabetes in Postmenopausal Women. *Diabetes Care* 37(3): 725-733.
- Ewing GW, Grakov IG (2013) A Further Review of the Genetic and Phenotypic Nature of Diabetes Mellitus. *Case Reports in Clinical Medicine* (9): 538-553.
- Ewing GW, Parvez SH (2010) The Multi-systemic Nature of Diabetes Mellitus: genotype or phenotype? *N Am J Med Sci* 2(10): 444-456.
- Porksen N, Hollingdal M, Juhl C, Butler P, Veldhuis JD, et al. (2002) Pulsatile insulin secretion: detection, regulation, and role in diabetes. *Diabetes* 51 (Suppl 1): S245-S254.
- English E, Idris I, Smith G, Dhataria K, Kilpatrick ES, et al. (2015) The Effect of anaemia and abnormalities of erythrocyte indices on HbA1c analysis: a systematic review. *Diabetologia* 58(7): 1409-1421.
- McDonald TJ, Warren R (2014) Diagnostic confusion? Repeat HbA1c for the diagnosis of diabetes. *Diabetes care* 37(6): e135-e136.
- Singh MM, Devi R, Saini V (2014) Using haemoglobin A_{1c} to diagnose type 2 diabetes or to identify people at high risk of diabetes. *BMJ* 348: g2867.
- Ewing GW (2008) NHS must make greater use of information technology. The quality of data - not the quantity. *BMJ* 337: a2303.
- Ewing GW, Parvez SH (2011) Mathematical Modeling the Systemic Regulation of Blood Glucose: 'a top-down' Systems Biology Approach. *NeuroEndocrine Letters* 32(4): 371-379.
- Ewing GW, Ewing EN (2008) NeuroRegulation of the Physiological Systems by the Autonomic Nervous System - their relationship to Insulin Resistance and Metabolic Syndrome. *Biogenic Amines* 22(4-5): 208-239.
- Ewing GW (2015) A Framework for a Mathematical Model of the Autonomic Nervous System and Physiological Systems using the NeuroRegulation of Blood Glucose as an Example. *J Comput Sci Syst Biol* 8(2): 59-73.
- Ewing GW (2016) Further Perspectives on Diabetes: Neuroregulation of Blood Glucose. *Neuroscience and Bio-medical Engineering (NBE)* 4(2): 75-83.
- Ewing GW, Ewing EN (2008) Cognition, the Autonomic Nervous System and the Physiological Systems. *Biogenic Amines* 22(3): 140-163.
- Ewing GW, Parvez SH, Grakov IG (2011) Further Observations on Visual Perception: the influence of pathologies upon the absorption of light and emission of bioluminescence. *The Open Systems Biology Journal* 4: 1-7.
- Ewing GW, Ewing EN (2009) Computer Diagnosis in Cardiology. *N Am J Med Sci* 1: 152-159.
- Ewing GW, Ewing EN, Parvez SH (2009) The Multi-systemic Origins of Migraine. *Biogenic Amines* 23(1): 1-52.
- Ewing GW (2015) Case Study: the Determination a Complex Multi-Systemic Medical Condition by a Cognitive, Virtual Scanning Technique. *Case Reports in Clinical Medicine* 4(6): 209-221.
- Ewing GW (2016) The Use of Strannik Virtual Scanning as a Modality for the Earliest Screening of the Pathological Correlates of Alzheimer's disease. *Human Frontier Science Program (HFSP) Journal* 10(2):2-20.
- Ewing G (2016) What is the function of the Brain? What does it do and how does it do it? It functions as a Neuroregulator, which continuously regulates the Autonomic Nervous System and Physiological Systems, and enables us to Recognise that Sleep Exhibits the Characteristics of a Neurally Regulated Physiological System. *J Neurol Psychol* 4(2): 9.
- Nwose EU, Ewing GW, Ewing EN (2009) Migraine can be managed with Virtual Scanning: case report. *The Open Complementary Medicine Journal* 1: 16-18.
- Ewing GW, Nwose EU, Ewing EN (2009) Obstructive Sleep Apnea Management with Interactive Computer Technology and Nutrition: Two Case Reports. *J Altern Complement Med* 15(12): 1379-1381.