Elevating the Human Spirit:

The Conflict Between Medical Statistics and Individuality During the Enlightenment

Marcello DiStasio

SUNY Downstate College of Medicine

November 2013

Introduction: The application of the scientific method to medicine

The study of medicine, health, and disease has enjoyed a long dialog with the natural sciences. In recent history, the alignment between allopathic medicine and scientific research-oriented universities has created a powerful system for discovering and disseminating information. The modern biomedical academic and private research system creates verifiable knowledge about physiology and pathology rooted in replicable experiments. In addition to (partly) dispelling charlatanism, this program of discovery drives progress by acquiring facts in a systematic way, and fitting them into evolving models of biological processes. By doing so, consistent recommendations for medical practice can be made by researchers and reviewers. Today's clinicians build their practice on a body of work that, while subject to imprecision, social pressure, and occasional revolution, is ultimately anchored to experiments and observational data. The challenge presented to the clinician is now finding a way to synthesize information from scientific studies involving large experimental populations or models into particular recommendations to patients.

Fortunately, the scientific method and the practice of statistical analysis allows users of scientific findings to compare and evaluate the reliability of findings and the certainty with which they can be applied under various circumstances. Summary statistics like, “positive predictive value”, “number needed to treat”, and “95% confidence interval”, are important tools to any physician trying to incorporate new knowledge into his or her practice or into the evaluation and plan made for an individual patient. I will discuss here some ways in which statistical science has influenced medicine, and how questions arising from health and public wellness concerns have driven advances in the practice of statistical analysis. Writing about the intersection of two domains of practical thought is difficult, since the cooperations and conflicts between them incorporate ideas, personalities, historical contingencies, and social utilization of each. Rather than attempt an exhaustive survey of the many important early instances of quantitative scientists investigating physiology or medicine, I have chosen a few examples that highlight what I consider to be the essential steps towards data-driven thinking in the practice of medicine and prevention of disease. These steps are:

- The understanding of systematic and random variation in humans and diseases
- The study of the effectiveness of interventions
- The application of scientific knowledge about populations to cases of individuals

I will focus on a couple key developments that took place during the 17th, 18th, and 19th centuries in Europe, an era that saw a notable acceleration in the use of numeric and statistical methods across the sciences. Medicine was no exception. Far from it; medical questions often drove and focused statistical thinking during this time.

Roles of science in medicine

The application of scientific method has benefited medical practice by systematizing observations about the functions of the human body and its response to intervention. Building this body of knowledge has created an enormous outgrowth of the roles for scientific investigation, including classification of pathophysiologies, identification of causative agents, identification of transmitting vectors, discovery of novel therapies, and establishment of diagnostic tests. All of these pursuits have immense impact on the practice of medicine and the health and well-being of the population, and all are rooted in (or at least accelerated by) the collection and analysis of quantitative data.

Modern medicine seems to be permeated by collections of data and conclusions drawn from these collections. Obviously this wasn't always the case. The relationship between science and medicine has been one of co-development, rather than revolution. In many ways, the advancement of the understanding of human biology and medical treatment paralleled the advancement in scientific investigation of the natural world. As the prevailing attitudes towards philosophy and reason changed from antiquity through the eras of reason and enlightenment and on through the modern era, so changed the attitudes surrounding health, disease, and the means of affecting them.

Many of the earliest applications of the modern scientific method were to problems in human disease and physiology. Advancements in medicine accompanied the progress in mathematics, physics, chemistry, astronomy, and biology that took place during the scientific revolution of the 16th, 17th, and 18th centuries. While Galileo was fighting the good fight for heliocentrism, William Harvey was working on a detailed study of the heart and blood, which culminated in the book *De Motu Cordis*, in which he overturned Galen's theory that the venous and arterial systems were separate dispersal systems, postulating instead the blood moves in a circuit. While

Isaac Newton was watching white light split into rainbows as it passed through a prism, Robert Hooke and his hand-crafted gold-inlaid microscope were busy formulating the cell theory of biology (all cells arise from other cells) and proposing that respiration involved the consumption of some component of the air. Edward Jenner developed the first vaccine, against smallpox, in 1790s Britain, becoming the "father of immunology" around the same time Allesandro Volta discovered electrochemical series and invented the battery. These discoveries helped spread the idea that systematic laws governing unseen elements could be deduced through experiments. When applied to human biology, experiment and analysis began increasing the sophisitication of medical knowledge as well.

Antiquity

Classical medicine was not without its rigor. An early example of what can be construed as a clinical trial appears in the Old Testament, in the story of Daniel, a court noble in Babylon. The first chapter of the book of Daniel includes the following passage:

> *Daniel resolved not to defile himself with the royal food and wine, and he asked the chief official for permission not to defile himself this way.*[9] *Now God had caused the official to show favor and compassion to Daniel,*[10] *but the official told Daniel, "I am afraid of my lord the king, who has assigned your food and drink. Why should he see you looking worse than the other young men your age? The king would then have my head because of you."*[11] *Daniel then said to the guard whom the chief official had appointed over Daniel, Hananiah, Mishael and Azariah,*[12] *"Please test your servants for ten days: Give us nothing but vegetables to eat and water to drink.*[13] *Then compare our appearance with that of the young men who eat the royal food, and treat your servants in accordance with what you see."*[14] *So he agreed to this and tested them for ten days.*[15] *At the end of the ten days they looked healthier and better nourished than any of the young men who ate the royal food.*[16] *So the guard took away their choice food and the wine they were to drink and gave them vegetables instead.*

By making comparisons of outcomes in an experimental group (the servants) with those of a control group (the young royals), Daniel was able to intuit the superiority of the vegetable diet.

The empiricist tradition emphasized three important pillars of clinical experience: learn from one's own observations, learn from past and present colleagues, and reason by analogy with cases thus learned. The famous Roman physician Galen emphasized the union of reason, medical experience, and experimentation in formulating an understanding of illness. He studied and wrote extensively on philosophy, merging his ideas with those of the Empiricists, who focused on experimentation, and the Rationalists, who thought that new knowledge should be deduced logically from established theory. Some of Galen's ideas had a decidedly statistical bent to them. He wrote of the idea that random variations can confound understanding, so one should be careful to review many cases before coming to a conclusion. He also comments on the hope that conclusions about whole populations can be reached after observation of a representative sample, but there is danger in excessive extrapolation. He also demonstrated remarkable insight into the fact that health and disease represent a continuum, along which quantification provides a valuable guide to the need for intervention. Galen's work contained some valuable pearls of scientific thinking, but was so widely adopted as dogma that many of his fallacious ideas about physiology took hold for almost a thousand years, preventing much experimentation and advancement of understanding.

The First Steps: Quantifying Observations

Because human health and disease have such direct impact on the operations of society and the state, their analysis has enjoyed more complete recording and treatment during the early enlightenment period than most other natural subjects. In fact, a work cited often as the first descriptive statistical analysis is John Graunt's 1662 book Natural and Political Observations made upon the Bills of Mortality, which was formulated as part of an attempt on the part of King Charles II to predict the spread of Bubonic plague in London. Basing his observations on various demographic findings and burial records, Graunt was one of the first to systematically analyze census data and use it to draw conclusions about the initiation, progression, and spread of disease [Graunt, 1662].

Bills of Mortality were public postings of the causes of death of the citizenry of London, as collected by parish clerks starting in the late 16th century. Weekly postings were begun after an outbreak of plague in 1603, and continued for over 200 years. A perusal of the causes of death yields some interesting cases including falling into a hot copper vat in a gin distillery (in 1728, at the height of the 18th century British "gin craze"), goring by a bull, and falls, through thin pond ice or from height. More common are the many psuedo-medical descriptions made by untrained record keepers, including King's evil (scrofula, thought to be cured by the touch of the monarch), Quinsy (tonsilitis), Falling sickness (epilepsy), Tympany (bowel obstruction), Horseshoehead (encephalitis), Rising of the Lights (croup), Livergrown, Looseness, Stranguary, Teeth, Worms, and Purples. Graunt, in a departure from his usual work as a successful haberdasher and warden, tabulated these cases and made a report covering the years between 1629 and 1659. Though he was elected to the Royal Society on its merits (as well as the directive of Charles II, against those who would deny a shopkeeper admission), he died penniless, having lost all his possession in the Great Fire of London, and all his credibility after converting to Catholicism (whose ranks were accused of setting the blaze). At least one copy of his Observations apparently survived the fire that consumed some 13,500 houses, and stands today as the first known example of a mortality table, an instrument widely used in insurance, epidemiology, and biostatistics. Graunt used his table to infer the probability of death due to various causes, including old age, and noted the difference between chronic diseases which claimed a constant proportion of lives each year and acute illnesses whose tolls fluctuated from year to year. In so doing, Graunt was making a very innovative foray into the practice of statistical inference. Graunt's method of choice was simple averaging, since other concepts that eventually became central to the study of populations (such as probability) were yet to be invented. Nonetheless, Graunt was able to achieve his objective of finding stable laws that predicted causes of death in the population [Jones, 1945]. This was a significant advance; he had used data to quantify tendencies and thus infer statistical laws.

The Development of Statistics

Mathematical statistics is a branch of applied math that employs the concepts of probability theory and statistical inference. These theories provide the underpinnings for statistical practice, to which the unqualified word "statistics" usually refers. A reasonable modern definition of statistics is "the science of collecting, organizing, interpreting, and reporting data, where the data consists of observations taken in the real world". [Fienberg, 1992][Porter, 1986]. The word statistics first appeared in the 18th century, referring to the collection of data about the state, meaning governmental and administrative data. Great advances in the understanding of probability and chance were made by mathematicians like Poisson, de Moivre, and Laplace during this time, and they applied them to the derivation of statistical laws from collections of data. Spurred by the advances made in physics happening in the same period, confidence was high that theories were capable of explaining natural phenomena, and that hidden natural laws governed complex systems.

One important natural law is the tendency for random events to distribute themselves in predictable patterns. The idea that using multiple data points to capture a central parameter of this distribution is fundamental to the practice of statistics. As an important example, Abraham de Moivre, a mathematician often called upon by gamblers to make recommendations, analyzed binomial outcomes (i.e. "go vs. no-go" trials such as coin flips) in his book *The Doctrine Of Chances*, published in 1718 [Schneider, 2005]. He took up the question "If you flip a fair coin a hundred times, what is the probability of getting 70 or more heads?" He plotted the number of heads outcomes in many of such series, and found that as more and more series were added, an increasingly smooth curve emerged. He knew that if he could find a formula for this curve, he could predict the probability of these outcomes ahead of time. Technically, this process constitutes a binomial distribution, but de Moivre discovered a curve that very closely approximates the binomial distribution and has even wider uses: the normal distribution. Later it was shown that for very large samples, the normal distribution becomes a better and better approximation to the binomial distribution. The normal distribution and its characteristic bell-shaped curve were used extensively by Carl Friedrich Gauss in his analysis of errors in the measurements of the locations of astronomical objects [Gauss, 1857], and has come to be commonly called the 'Gaussian distribution'.

The normal distribution arises in a number of important situations. Analyzing errors, as Gauss did, is one of the most frequently performed procedures in statistics. Measurement errors arise because of two processes: systematic bias in the measuring equipment or conditions, and random fluctuations in the equipment or conditions. The random fluctuations are basically the result of tiny random occurrences at every step along the measurement process. The total error is the sum of all these small component errors. This total error is normally distributed. In fact, it has to be, assuming a few simple conditions are true (and they often are). This is the conclusion of the Central Limit Theorem, which was discovered by Laplace soon after de Moivre published his analysis of games of chance. The Central Limit Theorem states that if groups of random numbers are taken independently from the same distribution and added, the sums will be distributed according the normal distribution. This means that by following the normal distribution equation, we are able to predict the probability of finding sums (and therefore averages) that are above or below any given value. By characterizing the behavior of errors with an equation, the early statisticians gave science a powerful tool for uncovering the true laws of nature as revealed by data. The variations in observed data can now be understood as a combination of systematic influences and randomness. Quantifying and characterizing the random components gives a clearer picture of systematic components. Randomness cannot be eliminated, but with equations to describe its aggregate effects, we are able to establish how much it contaminates our measurements of systematic effects. This separation into signal and noise is a major theoretical concept that has enabled countless scientific developments.

The Study of Disease (Lumpers v. Splitters)

Diseases, human characteristics, and almost any other collection of data, can be organized according to two modes of thought: lumping and splitting (a practice commonly debated in taxonomy). Lumping emphasizes the similarities between individuals, and looks to fit new entities into known categories. Splitting emphasizes differences between individuals, and seeks to fit or refine the classification scheme to accommodate the characteristics of new entities. Probability theory divides along similar lines. Bernoulli, or 'frequentist' approaches look at the data observed as random, and guess how likely it is for this data to occur given a fixed hypothesis. That is, they look for $\mathrm{P}(D|H)$. The Bayesian approach treats the data as fixed (since it's all the knowledge available), and asks how likely any particular hypothesis is (among all alternatives). Thus, when a study is performed by a frequentist, they are postulating a hypothesis (e.g. lung cancer incidence is higher among cigarette smokers), and evaluating how likely it would be to observe the data they see (e.g. 20% of smokers get lung cancer and 5% of non-smokers get lung cancer). If the data is more likely given the hypothesis versus given the null hypothesis (e.g. there is no difference), then they conclude the hypothesis is true. Thus, the analysis lumps data together based on similarities, and asks whether the resulting groups are consistent with the hypothesis under consideration. The Bayesian tries to evaluate the probability of a hypothesis being true (e.g. the moon is made of cheese) given the available data about it (e.g. it looks like cheese through a telescope, but we found rocks when we visited). That is, the Bayesian evaluates $\mathrm{P}(H|D)$. The nature of the moon is fixed; it either is made of cheese or it isn't. But the Bayesian scholar is trying to establish his certainty about the hypothesis. The Bayesian approach tries to quantify the subjective level of belief about one or more hypotheses. In other words, it leaves data points split apart from one another, and tries to find the hypothesis most consistent with all the data as it is.

This difference in philosophy underlies the dilemma faced by medical researchers and clinicians. When evaluating a single patient, the clinician is establishing a level of certainty about hypotheses given only the data about that one patient. The physician's job is to make a plan for care based on estimated likelihoods of diagnoses, responses to treatments, and prognoses in that one patient. All the physician has to go on is the data on the patient's condition acquired by personal histories, physical exams, and laboratory studies performed on that single patient. Meanwhile, epidemiological studies provide data about populations, in which tests and treatments are applied over and over, to many subjects. This corresponds more closely to the frequentist interpretation of probability. Thus, to make informed decisions, the treating physician has to decide whether to "lump" the current patient in with those subjects used in the study, or to focus on differences specific to this individual, and "split" him from the group. In a patient-centered model of medical care, all dimensions of the patient's condition must be considered; some data about the patient will be lumpable, and some must be split. Today's doctors are responsible for bridging both types of probability theory (even if just implicitly), as they incorporate knowledge from clinical trials (and other scientific studies) into their care of each individual person.

An early subscriber to the lumping school of thought was Adolphe Quetelet. Quetelet was a precocious student of mathematics (and of painting and opera) in French Period Flemish Belgium [Eknoyan, 2008]. As a young professor of mathematics, after taking a position at the Athene in Brussels, Quetelet got the chance to visit Paris, where he met the central planners of modern statistical theory, including Fourier, Poisson, and Laplace. Each of these men has entire branches of modern math emblazoned with his eponym, an honor never bestowed on Quetelet. In historically-conscious publications, however, the term "Quetelet Index" gracefully stands in for the widely fretted-over Body Mass Index (BMI). BMI has endured as an important statistic summarizing body composition during the two hundred years since Adolphe introduced as part of his effort to characterize the "average man". BMI is a simple numeric index (mass/height2) for a person's body shape. While not applicable to all humans (infants and the very tall are better represented with modified indexes derived conceptually from BMI), BMI itself is still widely used in stratifying subjects in medical studies and in advising individual patients about their obesity-related risks using an objective criteria. The latter highlights an important social use of summarizing statistics. A quantitative scale allows doctors, nutritionists, and other providers to tell a patient that he “has a BMI of 30” rather than that he “is overweight”, “Pickwickian”, or another euphemism for “fat”. Not only is this phrase less judgmental, it is precise; the patient is less likely to go home, examine himself in the bathroom mirror, and decide that his godlike physique couldn’t possibly be contributing to his health risks. The Quetelet index is essentially an attempt to distill many subtle differences in body composition into a summary statistic. Quetelet took data on height and weight from many subjects and, following a frequentist or lumping mentality, he hypothesized that there is an central tendency to human height and weight and that they have a geometric relationship, and then he showed that the data he observed were highly likely given that hypothesis.

Case Study: James Lind and the Prevention of Scurvy

The English Channel, which separates Britain from France and the rest of mainland Europe, was formed by torrential flooding about 200,000 years ago. It provides a rich fishing ground and an important means of crossing into and out of Britain, enabling migration and trade, but, just as importantly, it forms a barrier during times of conflict. In the 1700s, Britain and Spain were engaged a series of disputes, many over trading and territory rights in the colonies. Britain had amassed a powerful maritime presence, with around fifty thousand men in the Royal Navy and merchant marines in peacetime, and more than double that number during wars. Many of these sailors had been impressed into service, conscripted by the royal “Press-gangs” with the threat of hanging if they tried to desert. The conditions aboard ships were unsavory and brutish. Drunkenness was the most commonly cited reason for floggings [Jamieson, 1999], and fights were frequent.

Ships’ rations were controlled by the state Victualing Board, and each ship had a purser whose job it was to supply food and other goods like candles and coal. A typical food allowance on a Royal Navy ship included biscuits, beer, dried salted beef and pork, peas, oatmeal, sugar, cheese, and butter. There was a stark absence of fruits and vegetables. Scurvy was a prevalent threat at sea. Vitamin C, the micronutrient whose absence causes the disease, is required for the synthesis of collagen, the major component of connective tissues. Collagen can be found in skin, tendons, ligaments, the cornea, blood vessels, cartilage, and the endomysial component of muscles. Collagen is also an essential requirement for effective wound healing. When people get insufficient vitamin C in their diet, the collagen produced by fibroblasts lack the stabilizing crosslinks that give it its “springiness”, and wound healing is impaired. Other symptoms of scurvy include fatigue, malaise, bleeding gums, loss of teeth, fever, and neuropathy, eventually progressing to death. Reports of the toll of scurvy on the health of sailors must be evaluated cautiously, since the term was often applied to any condition doctors could not identify, but it is clear that it was a frequent cause of mortality. The disease, called scorbitus in Latin, has been known since antiquity, and is referenced by Hippocrates [Carpenter, 1986]. The modern word ascorbic acid, the proper name for vitamin C, has its origin in the Latin term. The ability to synthesize vitamin C was lost in the phylogenetic clade that contains humans approximately 60 million years ago, presumably because our ancestors were able to obtain the vitamin in such great quantities in their diet that the gene to produce it was under very little pressure to persist [Drouin et al., 2011]. Other species that are unable to synthesize it, and must get it from a diet of leaves, fruits, and vegetables, include bats, some birds, and teleost fishes.

Just before the War of Austrian Succession (1740-1748), a young physician named James Lind joined the British navy as a surgeons' mate. During the war, he transferred to the HMS Salisbury, under the command of one Captain George Edgecumbe during the Admiralty of Sir George Peacock. The Salisbury was a fourth-rate warship carrying 50 cannons that fired cannonballs in the 12-32lb range, with the larger ones capable of crippling a wooden enemy ship. It was a formidable ship, created along with its sisters out of the drive to outclass and outsize the navies of Spain and France.

The ship was engaged in patrolling the English Channel and the Bay of Biscay, a much-feared triangular expanse of open Atlantic water situated between Spain and France. As the Gulf Stream hits the continental shelf that divides the bay it breaks up, forming complex circular currents and high waves. The attendant storms and high sea conditions make crossing the bay perilous. Many ships have been turned around, lost, or wrecked in sudden gales. In presumably calmer conditions during the spring of 1747, Lind performed the experiment that earned him a place in the pantheon of medical experimentalists, administering various treatments to a group of stricken sailors. The Salisbury had departed Portsmouth two months earlier, captured a French privateer, returned the spoils to the port of Plymouth, and returned to sea, quickly capturing a French fishing boat and begin a patrolling mission in Biscay, around the estuary that supplies the rich vineyards of the Loire valley. On May 20, Lind selected twelve men suffering from scurvy and assigned two each to receive treatments of oranges, lemons, sea water, vinegar, elixir of vitriol (a mixture of sulfuric acid, alcohol, and ginger or cinnamon), or a purgative mixture of garlic, mustard seed, tamarind, and other spices mixed into barley water. He returned to medical school in Edinburgh to complete his MD degree and published his findings in a book titled *A treatise of the scurvy* [Lind, 1753]. In it, he describes his experiment as follows:

> *On the 20th of May, 1747, I took twelve patients in the scurvy, on board the Salisbury at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of the knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; and had one diet common to all, viz. water-gruel sweetened with sugar in the morning; fresh mutton-broth often times for dinner; at other times puddings, boiled biscuit with sugar etc. and for supper, barley and raisins, rice and currants, sago*

> *and wine, or the like. Two of these were ordered each a quart of cyder a day. Two others took twenty five gutts of elixir vitriol three times a day, upon an empty stomach; using a gargle strongly acidulated with it for their mouths. Two others took two spoonfuls of vinegar three times a day upon an empty stomach; having their gruels and their other food well acidulated with it, as also the gargle for their mouth. Two of the worst patients, with the tendons in the L arm rigid, (a symptom none of the rest had) were put under a course of sea-water. Of this they drank half a pint every day, and sometimes more or less as it operated, by way of gentle physic. Two others had each two oranges and one lemon given them every day. These they eat with greediness at different times, upon an empty stomach. They continued but six days under this course, having consumed the quantity that could be spared. The two remaining patients took the bigness of a nutmeg three times a day, of an electuary recommended by an hospital-surgeon, made of garlic, mustard seed, rad. raphan. Balsam of Peru, and gum myrrh; using for common drink, barley water well acidulated with tamarinds; by a decoction of which, with the addition of cremor tartar, they were gently purged three or four times during the course. The consequence was, that the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them, being at the end of six days fit for duty. The spots were not indeed at that time quite off his body, nor his gums sound; but without any other medicine, than a gargarism of elixir vitriol, he became quite healthy before we came into Plymouth which was on the 16th of June. The other was the best recovered of any in his condition; and being now deemed pretty well, was appointed nurse to the rest of the sick. [Lind, 1753]*

Lind's report shows a number of important experimental principles at work. He had chosen subjects that were "as similar as [he] could have them," being housed in the same conditions and eating the same diet. This represents an attempt to satisfy scientific principle of Descartes that one should eliminate all but one of the possible relationships between the items under investigation, though there is no evidence that Lind had premeditated on the scientific method. Nonetheless, Lind's book represented an insightful piece of work, grounded solidly in the historical knowledge of scurvy but bringing a new tool to the study: the use of a "fair test". He avoided selection bias, and tried to minimize confounding factors. The study falls short of modern standards for clinical trials in some aspects. He does not address his allocation criteria (today's statisticians urge randomness), and his numbers are obviously quite small. Perhaps more importantly, Lind does not draw the concrete conclusion that citrus fruit, or some component of it, is the cure for scurvy. The closest he comes to a definitive statement about the role of citrus fruits in the cure of scurvy is the following passage:

> *Some new preservative against the scurvy might in this treatise have*

> *been recommended; several indeed might have been proposed, and with great show of probability of their success; and their novelty might perhaps have procured them a favourable reception in the world. But these fruits have this peculiar advantage above anything that can be proposed for trial, that their experienced virtues have stood the test of nearly 200 years. [Lind, 1753]*

Lind, influenced by the prevailing views of scurvy in his day, misinterpreted the effect of citrus fruits to be a factor in digestion, aiding in the absorption of nutrients that improved the conditioning of patients that had been depressed by poor diet, stale air, and confinement. Nonetheless, the strength of his method produced a valuable identification of a true effect. Another tenuous aspect of Lind's study is in the selection of treatments. One may speculate about what the outcome would have been should Lind have used food items or compounds that all lacked vitamin C, or items that all contained it. No true differential effect would have been present. This highlights a key point about clinical trials, and scientific experiment in general: all hypotheses are tested by comparison. The use of controls is of paramount importance. If no comparison is made (or a faulty comparison is made between elements that do not differ), then the exercise is no more than educated empiricism. Fortunately for Lind and sailors everywhere, the treatments he compared truly differed in the feature under investigation. Lind's selections were not without basis, however. He was extremely well versed in the theories and treatments of his day, and the available empiric or "folk" remedies provided a starting point for making informed selections. The idea that citrus fruit is an effective treatment for scurvy had existed for a long time. Even Vasco de Gama, in 1498, made a note that he had sent a man to shore to "bring a supply of oranges that were much desired by our sick." In 1593, while sailing around the southern shores of Brazil, the pirate knight Richard Hawkins wrote "There was great joy amongst my company and many with the sight of the oranges and lemons seemed to recover heart. This is a wonderful secret of the power and wisdom of God that hath hidden so great and unknown virtue in this fruit to be a certain remedy for this infirmity." James Lind did not mention these examples in the quite extensive review of the medical approach to scurvy he provided in the introduction to his Treatise, but he does acknowledge number of his predecessors who had written on the antiscorbutic effect of citrus fruits. This type knowledge has immense value to science. It provides a background of prior information upon which comparative hypotheses must be based. As scientific trials build upon this knowledge, confirming some ideas and discarding others, the errors in empiric knowledge are stripped away, and new more complex hypotheses can be formed on solid footing. Unbiased comparisons like Lind's provide data. Interpretations may change, but properly acquired data do not. Future studies corroborated Lind's findings because the data he generated (i.e. the fact that scurvy responds favorably to citrus) was properly acquired.

The interpretation of data, however, is subjective and thus susceptible to practical and social constraints. Lind's findings were not widely applied upon their publication. This is partly due to Lind's own reticence. He was reluctant to recommend a treatment whose mechanism of action he did not understand. Medical practice at the time was dominated by theories; largely untested, but logical and somewhat internally consistent. Induction of new knowledge into a system that could not account for it was an anathema to many physicians, even when faced with convincing data. Lind's data did not gain immediate traction.

Lind was most definitely aware of the dichotomy between theory and experiment. He states clearly ". . . I shall propose nothing dictated merely from theory; but shall confirm all by experience and facts, the surest and most unerring guides." His thinking was sound, but perhaps his conviction was weak; he did only make use of a dozen subjects in his now famous experiment, after all. In his writings, he made very few decisive statements about the singular efficacy of citrus, and later provided a recipe for a inspissation of orange juice that has since been shown to contain very little vitamin C after a month's storage [Hughes, 1975]. The admiralty of the British Navy did not require ships to carry fresh until 40 years after Lind's work was published, at the urging the court physician Gilbert Blaine. These things so not diminish Lind's contribution to modern medicine. It has been described as "the first deliberately planned therapeutic trial" [Thomas, 1969], and was certainly one of the earliest, and clearly one that made its way most forcefully into the medical cannon. Lind was never exactly able to do what he set out do: In his own words, "to remove a great deal of rubbish...before the subject could be set in clear and proper light,". The theories of Lind's day didn't permit it. Lind's work preceded a clear understanding of the pathophysiology of scurvy, and didn't benefit from knowledge of the existence and role of micronutrients (which only came after the animal experiments of Frederick Hopkins and Christiaan Eijkman in the early 20th century). Though his interpretation lacked the context of modern nutritional science, and adoption of his recommendations lagged considerably, his data proved more powerful than theory or social circumstance, and the idea eventually caught up.

On dry land, Lind continued his treatments, giving lemon juice to many patients at Haslar hospital in Portsmouth, and updated his book accordingly, commenting on the success of the treatment. He provided an extensive review of what was then known about scurvy, and included critical reviews of over fifty medical texts. He published a few more books, updating his guide for preventing and treating scorbutic illness, and making extensive other recommendations for the preservation of the the health of sailors, making him one of the first authors in occupational health.

Case Study: Pierre Louis and the Statistics of Blood-letting

The practice of venisection and phlebotomy grew out the Hippocratic humoral tradition, which held that disease was a state of imbalance among the principal body fluids. Hippocrates believed that the function of menstruation was to eliminate bad humors from women, and that by extension, the release of bad humors from the sick could restore the equilibrium of vital fluids and forces necessary for health. The practice was expanded and codified in Rome, fueled largely by the prolific writings of Galen, who had been educated in the Alexandrian school to believe that many diseases were the result of plethoras, or the overabundance of bodily fluids. When Galen discovered that veins transmitted blood (not air) along with arteries, it reinforced the practice, based on a a further misinterpretation of the function of blood. Galen and his contemporaries believed that blood was created centrally (particularly in the liver) and flowed in one direction to the periphery where it was consumed. This raised the possibility of unused blood stagnating in the periphery and inducing bodily dysfunction. The solution was thus to purge the body of this noxious fluid.

Galen was a prolific and forceful advocate for his ideas, and produced a complicated system describing the amount of blood to be let based on the patient's age, constitution, ailment, geographic location, and on the weather. His ideas took hold and the practice persisted as a mainstay of medical treatment until the 19th century. Therapeutic phlebotomy was, in fact, the most commonly used treatment in ancient medicine. Illnesses causing the skin to become red (particularly fevers and rashes), were thought best treated by blood letting. Religious texts even carried instructions governing the appropriate day and setting for blood letting. Leeches, whose saliva carries anticoagulant proteins (particularly hirudin, a potent inhibitor of thrombin, along with a number of anti-platelet aggregation factors), were widely used in Europe for therapeutic blood letting, a fact highlighted by the name of the most widely used species: Hirudo medicinalis, 42 million of which were imported into France for medicinal use in a single year. Franois Joseph Victor Broussais, an influential Parisian doctor in the early 1800s, held that most illnesses were due to inflammation in specific organs, and leeches were thus to be applied to areas of the body surface corresponding to the afflicted organ. Like many aspects of ancient medicine, the wide use pf phlebotomy had persisted because its effects are complex and subtle enough that effects on health were not immediately apparent to patients or physicians. Thus, when health improved phlebotomy was credited. When it declined, the illness was blamed.

Under our current understanding of hematology, blood letting has almost no role in curing disease, being deployed in only very particular circumstances (the classic examples being polycythemia vera, a condition in which red blood cells are produced too rapidly and excess iron is deposited in the tissues, and hereditary hemochromatosis, in which there is a pathologically high rate of absorption of iron from the gastrointestinal tract). We now know that excess blood loss can result in anemia, or insufficient hemoglobin in the blood, with its associated poor oxygenation of the tissues. So how did this practice, the mainstay of ancient medicine, fall to it's present (and deserved) disuse? The answer lies in an experiment performed by the inventor of clinical statistics, Pierre Charles Alexandre Louis.

Louis was born in 1787 in the Champagne region of France. His father was a wine merchant. After medical school and an interlude practicing in Russia, Louis took a job in Paris as a medical clerk at La Charitê Hospital. Acting as something akin to today's pathologists, he took notes on medically relevant facts, treatments rendered, clinical outcomes, and autopsy findings. Louis held doubts about Broussais' system of explaining disease, and published a number of arguments against it.

Louis selected a series of 77 cases of pneumonia diagnosis in patients with no known preexisting illness. He estimated the time of onset of illness in each patient as accurately as possible, and made a record of whether blood letting was performed early in the illness (less than 4 days after onset) or late in the course of illness (5-9 since onset). He made an attempt to select homogeneous sets of patients for each group, noting that the average age was the same for both, and all had a clinically similar form of pneumonia. He did note that the early blood letting group was somewhat enriched for patients over 50 years old as compared with the late group, and commented on the possible biasing effect this might have on the analysis. This study design represented one of the first attempts at a "retrospective case-control" structure. A case-control series is an observational study in which the efficacy of two treatments is compared by selecting a group of cases that are maximally similar, ideally differing only by the treatment applied. It remains an important tool in epidemiology today; in fact this study structure was the one employed in one of the landmark studies linking tobacco use to lung cancer by Doll and Hill in 1950.

Louis understood the necessity of controls in group comparison, stating as the aim of his work "...to ascertain whether, other things being equal, the patients who were bled on the first, second, third or fourth day of their illness, recovered more readily than those bled at a later period." He referred to this line of investigation as the "numerical method," and drew heavily on the ideas of Laplace and Marquis de Condorcet and their work on probability theory. He found that the duration of illness was on average 3 days shorter in patients bled early versus those bled late, but that more patients who had been bled early died (44% vs 25%). Louis' analysis led him to the conclusion that "the study of the general and local symptoms, the mortality and variations in the mean duration of pneumonitis, according to the period at which bloodletting was instituted; all establish narrow limits to the utility of this mode of treatment." Louis had demonstrated empirically that outcomes of phlebotomy treatment did not conform to the theory passed down to Broussais from antiquity. Apparently, the practice was already falling out of favor by the time Louis' study reached a large audience, but his work provided a convincing nail in the coffin. It is important to note that even though William Harvey's arguments for the circulation of blood and other more modern theories of the function of blood, based on experimental physiology, were well appreciated in Louis' day, clinical practice remained tied to an ancient system. This is a common (and not altogether improper) practice in medicine: experimental evidence builds theories while clinical epidemiology based on those theories alters practice.

Louis' legacy is more than the simple refutation of a thousand-year old piece of medical quackery. His methods for establishing the truth in the face of clinical uncertainty formed the basis for the modern case-control study, and have had a lasting effect on the way clinical knowledge is acquired and tested. Louis' study of blood letting carried him to prominence and a position as an instructor at Pitié-Salptrire Hospital in Paris, where he taught his methods to students who had traveled from all over the world to study medicine in the City of Lights. Among his students one can count Oliver Wendell Holmes Sr., a prominent Boston physician, and Drs. William Gerhard and Caspar Pennock, who returned to their native America and applied the numerical method to the study of infectious disease in Philadelphia, describing for the first time the difference between typhus and typhoid fever. Since the early 19th century, the use of the clinical trial has amplified and increased in sophistication, but remains rooted in the probability theory, group comparisons, and population-based thinking that Louis pioneered. He had carried out his research in the days before epidemiology was a formal discipline, but followed a decidedly modern program of data collection and analysis, and had a major impact on future study designs.

The controlled clinical trial was adopted by many important investigators in the 19th century. In the early 1830s, Jean Civiale compared mortality rates between groups of patients treated for bladder stones by two methods, traditional open lithotomy, and a new transurethal lithotripsy method he had developed, and found a lower mortality in patients subjected to the new method (2.2% vs. 18.8%). Joseph Lister reported average mortality rates before and after incorporation of his techniques for antisepsis during surgery in 1866. Louis Pasteur conducted a trial on sheep in 1881 that had all the features of a clinical trial using the numerical method and provided convincing evidence of the effectiveness of his vaccine against anthrax. And of course many modern randomized clinical trials have contributed to practice guidelines throughout medicine today. The first randomized clinical trial was undertaken at the English Medical Research Council (MRC) in 1948 to show the efficicay of streptomycin in treating pulmonary tuberculosis, and thousands more have been run since [Streptomycin in Tuberculosis Trials Committee, 1948][Stolberg et al., 2004].

Opposition and the primacy of the individual: The 1835 Paris Academy Commission

Even as statistical analysis and clinical trial methodology expanded, there were those who rightly retained focus on the individual patient. Some, however, were unable to reconcile the study of populations and experimental models with the art of healing the individual. One such man was Francois Double (1776-1842), a physician member of the the Paris Academy of Sciences and co-founder of the Acadmie Nationale de Mdecine. In 1835, the Academy convened a commission tasked with evaluating the application of the numerical method to medicine, and asked Double to act as reporter. The commission was a response to the recent publication by Civiale his finding that lithotripsy, a bloodless procedure for removing bladder stones, was superior to the traditional open surgical method of lithotomy. He came to this conclusion after collecting statistical data about outcomes from across Europe, building on Louis' methodological foundation. The Paris Academy commission, which included Simon Denis Poisson used Civiale's publication as an impetus to take a larger view of statistical methods in medicine. The commission was dubious (even Poisson, whose name has become so linked with statistical methods that he has an exponential distribution named after him), and ultimately critical of the application of the "calculus of probabilities" to medicine. They saw statistical reasoning as a fashionable trend, one of many that had purported to revolutionize medicine over the years. More importantly, they struck at the heart of the debate that continues today: How much of medical practice should follow preconceived recipes based on scientific findings, and how much of it should be logically deduced for an individual patient based on the unique aspects of his or her case. Double recorded this conflict adroitly:

> *In statistical matters...the first care before all else is to ignore that a man is an isolated individual and only to consider him as a fraction of the species. It is necessary to strip him of his individuality in order to eliminate any accidental qualities from the question. In applied medicine, on the contrary, the problem is always individual, the facts only presenting themselves one at a time.., and finally it is only a single man with all his idiosyncrasies that the doctor must treat. The masses remain completely out of the question [Poisson et al., 1835].*

Double also raised the concern about the medical field's ability to accrue of sufficient numbers of cases to permit probabilistic inference. He rightly followed this with the question of just how large these numbers of cases need to be. Though the modern theory of statistics gives us a number of ways of calculating numbers needed for given levels of statistical power, Double did not have access to these, and refused to concede to the numeric method any adequate level of certainty. The commission's report resonated widely with physicians who lacked the mathematical background to see how quantitative practice might be expanded to deal with the subtleties of human studies. Louis and his allies were not deterred, and formed their own commission to issue a rebuttal. They continued to advance the argument that with clarity and rigor, medicine could be transformed into a true science, while Double and his colleagues continued to assert that it was an error to "elevate the human spirit to that mathematical certainty found only in astronomy."

The Conflict Between Patient-Centered and Evidence-Based Medicine

There exists a fundamental conflict in medical science between the need to approach each patient as an individual and the abstract understanding of disease processes. Medicine is inherently a humanist enterprise. Doctors are concerned with patients as persons, each with his or her own needs, challenges, particular physiology, and perspective on disease. We trust our healthcare providers to consider our needs on a one-by-one basis, and often resent implications that we are interchangeable units being provided standardized care. We ruffle at the idea that governmental or insurance agencies might fix our level of care based on numerical cutoffs or formulas. It is scary to think that a peculiarity of our own ailment might be missed by a "plug-and-chug" system in which a rote computation determines the best courses of action. Since antiquity, medicine has been recognized as an art in which the practitioners instinct and judgment are central to care. Even in the era of modern technology our definition of wellness goes beyond achieving normal physiology to include patients' understanding of their body function and how it fits with their goals, resources, and daily activities. These things can only be appreciated on a case-by-case basis, and current health care models demand that decisions about these issues be shared between patient and caregiver.

There is a problem here. While medicine is indeed a humanist enterprise, the advancement of scientific understanding can only occur when we strip away individual details in order to look squarely at fundamental causes and effects. To do so is to ignore the specialness of each person, and tally only the facts that can be meaningfully compared to other cases in a database. Scientific clarity comes when confounding variables are discarded and simplified (but accurate) models are revealed. "To do science," wrote the great evolutionary biologist Robert MacArthur, "is to search for repeated patterns, not simply to accumulate facts." [MacArthur, 1972]

Medical science demands that the human organism must be studied in all its multifarious incarnations. We must survey the wide variations of normal and the wide variations of disease in order to contrast them meaningfully. The long succession of methodical physicians have, through the ages, recognized patterns and consistencies among the variety of illnesses encountered, forming the canon of medical knowledge housed in much-revered library stacks around the world. Much of this body has been built inductively, advancing from a collection of instances of disease to a systematic understanding.

This advancement in medicine has provided us with immense power to prevent, predict, diagnose, and treat disease, and to give an ever-widening proportion of humanity access to these abilities. It has largely paralleled the development of the scientific method in general. In spite of the massive benefits supplied by the scientific treatment of medicine, the practice has been slow to incorporate quantitative methods. This is not because of physician obstinacy or traditionalism. Rather, it reflects the fact that the medical enterprise is both a scientific exploration and an application. Most of the selfsame practitioners of data gathering, experimentation, and synthesis of knowledge are bound by their humanist ethical responsibility to their patients to treat them as individuals. This necessitates polymodal treatments, guesswork, and therapeutic decisions based on social and economic factors. These things can rarely be coerced to conform to the Cartesian principle that only the factor under investigation be allowed to vary. Modern multifactorial statistics has provided tools to deal with covariates, but they are not a substitute for good controls, and require large numbers of subjects.

Finally, studying human beings is hard. People have freedom and fallibility, pride and prejudice, tastes and traditions. It is completely impossible to discard all the confounding variation. Rather, the goal of the physician-researcher is to learn the principles that govern health and disease in the face of variation.

Enlightenment philosophy and mathematics paved the way for examining human conditions by the scientific method, but it was an early few individuals who shouldered the task of actually carrying out the first quantitative investigations. The theories of physiology with which they conceived their experiments were largely untested folk knowledge, and the theories of statistical reasoning that they brought to bear on their data were still in their nascent stages. Physicians like Louis and Lind, along with biometricians like Graunt and Quetelet, took important steps by using numerical methods in evaluating difficult questions about health. The central findings of their work, but not their interpretations, have stood the test of time. This is a testament to the fact that correctly acquired data can hold knowledge independent of the vagary of thought in those who acquire it. As the theory and technologies of statistics progress, they give us new tools with which to let the data “speak for itself”, eliminating human bias by requiring fewer assumptions about the data. In this way, the general knowledge grows, and the astute clinician makes use of this base when formulating recommendations for the individual.

References

[Best and Neuhauser, 2005] Best, M. and Neuhauser, D. (2005). Pierre charles alexandre louis: Master of the spirit of mathematical clinical science. *Qual Saf Health Care*, 14.

[Carpenter, 1986] Carpenter, K. (1986). *The History of Scurvy and Vitamin C.* Cambridge University Press. 9

[Chen, 2003] Chen, T. (2003). History of statistical thinking in medicine. In Liu, Y. and Fang, J., editors, *Advanced Medical Statistics*, chapter 1. World Scientific Publishing, Singapore.

[Drouin et al., 2011] Drouin, G., Godin, J., and Pagé, B. (2011). The genetics of vitamin c loss in vertebrates. *Curr. Genomics*, 12:371–378. doi: 10.2174/138920211796429736. 9

[Eknoyan, 2008] Eknoyan, G. (2008). Adolphe quetelet (1796-1874) - the average man and indices of obesity. *Nephrol. Dial. Transplant.*, 23:47–51. doi:10.1093/ndt/gfm517. 7

[Fienberg, 1992] Fienberg, S. (1992). A brief history of statistics in three and one-half chapters: A review essay. *Statistical Science*, 7:208–225. 5

[Galen, 1964] Galen (1964). *De anatomicis administrationibus libri ix. In: Kuhn CG. Claudii Galeni opera omni*, volume 2. Olms, Hildelsheim.

[Gauss, 1857] Gauss, C. (1857). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Arbientium. 1809. Hamburg. Perthes and Besser. English Translation: C Davis. Theory of the Motions of the Heavenly Bodies Moving About the Sun in Conic Sections*, volume Repr: 1963. Dover Publications. New York. 5

[Graunt, 1662] Graunt, J. (1662). *Observations On The Bills Of Mortality.* First edition. 4

[Hughes, 1975] Hughes, R. (1975). James lind and the cure of scurvy: an experimental approach. *Med Hist*, 19:342351. 12

[Jamieson, 1999] Jamieson, A. (1999). Tyranny of the lash? in the royal navy during the american war, 1776-1783. *The Northern Mariner/Le Marin du nord*, IX:53–66. 8

[Jones, 1945] Jones, H. (1945). John graunt and his bills of mortality. *Bull Med Libr Assoc*, 33:3–4. 5

[Lind, 1753] Lind, J. (1753). *A Treatise of the Scurvy in Three Parts. Containing an inquiry into the Nature, Causes and Cure of that Disease, together with a Critical and Chronological View of what has been published on the subject.* A. Millar, London. 9, 10, 11

[MacArthur, 1972] MacArthur, R. (1972). *Geographical Ecology: Patterns in the Distribution of Species.* Princeton University Press. 17

[Matthews, 2001] Matthews, J. (2001). Commentary: The paris academy of science report on jean civiales statistical research and the 19th century background to evidence-based medicine. *International Journal of Epidemiology*, 30:1249–1250.

[Morabia, 2006] Morabia, A. (2006). Pierre-charles-alexandre louis and the evaluations of bloodletting. *J R Soc Med*, 99:158–160.

[Murphy, 1981] Murphy, T. (1981). Medical knowledge and statistical methods in early nineteenth-century france. *Med Hist*, 25.

[Poisson et al., 1835] Poisson, S., Double, F., and et al (1835). Rapports: Recherches de statistique sur l'affection calculeuse, par m. le docteur civiale. *Comptes Rendus Hebdomadaires des Sances de l'Acadmie des Sciences*, 1:171–172. 16

[Porter, 1986] Porter, T. (1986). *The rise in Statistical Thinking, 1820-1900.* Princeton University Press. 5

[Schneider, 2005] Schneider, I. (2005). Abraham de moivre, the doctrine of chances (1718, 1738, 1756). 5

[Stolberg et al., 2004] Stolberg, H., Norman, G., and Trop, I. (2004). Randomized controlled trials. *Am J Roentgenol*, 183:153944. 15

[Streptomycin in Tuberculosis Trials Committee, 1948] Streptomycin in Tuberculosis Trials Committee (1948). Streptomycin treatment of pulmonary tuberculosis. a medical research council investigation. *Br Med J*, 2:76982. 15

[Thomas, 1969] Thomas, D. (1969). Experiment versus authority: James lind and benjamin rush. *New Eng J Med*, 281:932–933. 12