

Evolution and Application of Crowd Wisdom Techniques in Healthcare Decision Making : An Overview of Ethical and Moral Implications

Henry Asante Antwi^{1,2}, Zhou Lulin^{1,2} Samuel Owusu Mensah¹, Ethel Asante Antwi^{1,3}, Michael Owusu Akomeah^{1,4}

¹School of Management, Jiangsu University, 301 Xuefu Road, Zhenjiang, Jiangsu, P.R. China

²Institute of Medical Insurance and Healthcare Management, Jiangsu University, Zhenjiang, Jiangsu, P.R. China

³School of Graduate Studies, Ghana Technology University, Private Mail Bag, Abeka-Accra, Ghana

⁴Ghana Technology University Business School, Private Mail Bag, Abeka-Accra, Ghana

ABSTRACT

Identifying a clear path to reduce perturbation of expert opinion from truth and incorporating appropriate evidence based techniques to standardize its use remains an adventure in transition. This is the challenge with harnessing clinical experiences as crowd wisdom techniques for medical decision making. We attempt in our review, to examine the evolution of the crowd wisdom theories and the many ramifying models that have emanated from this theory. We identify and explain the various crowd wisdom models applied to the field of healthcare. We note several outstanding questions regarding the moral, ethical, legal and other factors that must be addressed and contextualized.

Keywords : Medical Decision Making, Magnum Opus, Crowd, Heyday Of Research, Tacit Knowledge, EMH, ANN, ACO, PSO

I. INTRODUCTION

Many notable scientific inventions are the product of academic curiosity for individuals' physical and mental qualities. Such is the origin of the Galton's "magnum opus" the wisdom of the crowd theory. Since then the wisdom of the crowd theory has been harnessed and optimized for several professional use and the health sector is no exception. We attempt in our review, to examine the evolution of the crowd wisdom theories and the many ramifying models that have emanated from this theory. We identify and explain the numerous crowd wisdom models applied to the field of healthcare. Following, we interrogate the possible legal, ethical and moral reasons that drive their contemporaneous use with evidence based models.

II. METHODS AND MATERIAL

A. Evolution of Crowd Wisdom

In 1906, scientist Francis Galton's curiosity for individuals' physical and mental qualities, in addition to his obsession for animal breeding, led him to become a seminal author of work on the 'wisdom of crowds' (Surowiecki, 2004). During what was originally intended as a leisurely day out for Galton at the annual West of England Fat Stock and Poultry Exhibition in Plymouth, he stumbled across a weight-judging competition where members of the public, skilled and unskilled alike in the task of judging the weight of a Fat Ox, were paying sixpence to guess the Ox's weight in the knowledge that the closest individual estimate to the actual weight of the Ox, once it had been "Slaughtered and Dressed", would win a prize (Galton, 1907). James Surowiecki in his book 'The Wisdom of Crowds' (2004) tells the story of Galton's decision to turn the competition into an 'impromptu' experiment. Galton's initial aim was to in fact affirm his belief that "the stupidity and wrong-headedness of many men and women being so great as

to be scarcely credible” (Quoted in Surowiecki, 2004). Yet Galton was to be surprised by his findings. He collated all of the 787 legible estimates and calculated the mean of these estimates, acquiring a figure of 1,197 pounds, one away from the correct weight of 1,198 pounds, an error of only 0.09% (Israeli & Silber, 2012). The underlying principles of these quite remarkable results are the foundations upon which this paper stands; the logic that “under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them” (Surowiecki, 2004). Around the time at which Galton published his findings, the traditional literature relating to collective judgements as opposed to those of the individual was somewhat to the contrary (Surowiecki, 2004). Charles Mackay, who published on the ‘Madness of Crowds’ in 1841, stated that “Men, it has been well said think in herds...It will be seen that they go mad in herds” whilst the speculator Bernard Baruch in the forward of the republication of Mackay’s work wrote “Anyone as taken as an individual is tolerably sensible and reasonable – as a member of a crowd, he at once becomes a blockhead”. Supporting the views of Mackay and Baruch were authors such as Thoreau (1962), Nietzsche (1866) and Carlyle (Quoted in Surowiecki, 2004) to name a few, yet perhaps the harshest critic of the wisdom of crowds was the French psychologist Gustave Le Bon in his 1895 publication ‘The Crowd : A Study of the Popular Mind’. Le Bon was an advocate for the belief that individual opinions are superior to those of the crowd and was also a ruthless critic of his antecedents such as Herbert Spencer (Le Bon, 1895). He utilised a chemical analogy to portray his standing that individuals collaborating in a crowd are like: “certain elements, combined to form a new body possessing properties quite different from those of the bodies that have served to form it” (Le Bon, 1895). Le Bon called the gathering of men, whether this related to traditional examples or rioters or rebels for instance, an “organised crowd” . He stated that “how much” an isolated individual “differs” from a crowd of which they are a part can be “easily measured”, yet he does not provide examples of measurements other than to declare that juries return verdict to which each of the individual jurors would disapprove whilst also deeming that “parliamentary assemblies adopt laws and measures of which each of their members would disapprove in his own person” (Le Bon, 1895).

Following the work of Le Bon and the counterarguments played out in Galton’s writings, the next major contributions within the field of group dynamics came during what Surowiecki describes as the “heyday of research” within the field (Surowiecki, 2004). Many of the experiments conducted during the period between 1920-1960 found a similar correlation to Galton. Many of these studies, however, used only a relatively small crowd. Kate Gordon (1924) carried out an experiment in which 200 students were asked to place in descending order a series of ten weights. In addition to taking a simple average of the accuracy of the students’ judgements, she also grouped the students “orders” or guesses into factions of five, ten and twenty, and then, as before, calculated correlations between the averages arising as a result of the groups. Her results comprehensively suggested that as the number participating within the crowd increases, so does the accuracy of their average estimate in addition to producing an overall crowd accuracy of 94%. This result could be perhaps paralleled with a mathematicians view in statistics that increasing a sample size would yield increased accuracy (Anderson et al, 2007). Following numerous similar studies (Bruce 1935; Gurnee 1937; Knight 1921; Shaw 1932), and corresponding similar outcomes, the power of aggregating information held by individuals amongst a crowd became more widely recognised. With such conclusive results being extracted from various studies the wisdom of the crowd phenomenon began to be acknowledged as more than a set of coincidental results.

In 1945, Friedrich Hayek, an Austrian economist and philosopher, published his paper ‘The Use of Knowledge in Society’. He noted a close corollary; that market prices aggregate dispersed knowledge about economic value (Hall, 2010). His work, written partly as a criticism of Joseph Schumpeter’s remarks on assumptions that individuals hold the knowledge on which they act in the financial marketplace (Schumpeter, 1942), was the first of its kind to apply in practice the aggregation of information research of his peers, in his case within the financial markets. He analysed the use of the price mechanism as a vehicle for aggregating “tacit knowledge” or information, writing: “The mere fact that there is one price for any commodity ... brings about the solution which ... might have been arrived at by one single mind possessing all the information which is, in fact, dispersed among all

the people involved in the process.” (Hayek, 1945) Surowiecki (2004) notes that despite his research, Hayek had a “fear of socialism and centralized authority” which led to him overestimating the difficulty of information aggregation and thus underestimating the potential benefits of such aggregation in markets. A modern example of this could be seen in pari-mutuel betting markets, a quite un-Hayekian occurrence due to one entity taking one side of all bets, yet one which is often highly accurate at forecasting future events within its odds framework (Hurley & McDonough, 1995). Despite Hayek’s apparent breakthrough, his work has been challenged since, particularly through the use of empirical studies (see: Boettke, Caceres & Martin 2012; Hurley and McDonough 1995; Smith 1982). Following the Hayek Hypothesis¹, the next major contribution to the literature came in 1970 in the shape of Fama’s Efficient Market Hypothesis (EMH). Pioneered by early work from Bachelier (1900) the EMH states that in an efficient market, future payoffs are correctly valued in the current period by the price at which they are trading, thereby making the current price the best predictor of an event occurring; an informationally efficient world (Rajakovich & Vladimirov 2009; Wolfers & Zitzewitz 2004). Thus a market is efficient if all available information is always fully reflected in the price (Foutz & Jank 2007). As Wolfers quite rightly suggests however, the effect to which the price mechanism can incorporate data and information is dependent upon whether this information is available at all to participants within the market (Wolfers, 2009). Categories of information exist, Weak, Semi-Strong and Strong², expressing the availability of information thus indicating how much information is represented by the price level (Fama, 1970; Vaughan Williams, 2005). The level to which information is included within a market price is known within modern day literature as Information Efficiency (Vaughan Williams, 2005).

When one reviews the literature elucidated above and reviews Hayek and Fama’s theories alongside such an example as Galton’s Ox experiment, there is perhaps, on one level, a realisation that Hayek and Fama may have discovered a key application of crowd wisdom. In a study conducted by Professor Jack Treynor in 1987, 56 students were asked to provide estimates of how many jelly beans were in a jar. The mean guess of the students was 871, representing a 97.6% level of accuracy, with only one of the 56 estimates getting

closer to the actual value of 850 (Treynor, 1987). In support of Treynor’s work, a similar study conducted by the researcher, again sampling estimations from 56 students showed a similar Both of these experiments would appear to affirm, in addition to other studies undertaken with respect to the dynamics of group decision making (see Dunning 2007; Hastie and Kameda 2005; Lorge et al. 1958; Sunstein 2006; Surowiecki 2004; Yaniv 2004), the ideal that independent judgments of a crowd of individuals (as measured by any form of central tendency) will be relatively accurate, even when most of the individuals in the crowd are ignorant and error prone (Surowiecki, 2004; Frederick et al, 2011).

B. Application of Crowd Wisdom Techniques in Healthcare

Since the ground-breaking work of Francis Galton, many different crowd wisdom techniques have been developed and applied in many different fields of endeavour including the healthcare sector. For example, in 1950, the Delphi method was developed by RAND for technology forecast but has since been used as a method to gather experience-based clinical knowledge. The results of these medical application of Delphi have been documented in studies such as by (Kurtzberg & Levanoni, 2002; Lausch, Schmidt, & Tischendorf, 2015; Sloane, 1997) all of whom studied its application in a clinical decision scenario while Sloane (1997) has applied it to a non-clinical medical scenario with reasonable degree of accuracy (albeit its weaknesses). With the popularisation of data mining techniques, many different forms of Swarm Intelligence (SI) and Artificial Neural Networks (ANN) have been experimented to support evidence based medicine in assessing (S. S. Jones et al., 2008; Biglarian et al., 2011; Spelt, Nilsson, Andersson, & Andersson, 2013); and predicting (Burke et al., 1997; Chen et al., 2007; Shi, Tsai, et al., 2012) more complex biological systems and medical scenarios with greater degree of accuracy over the conventional statistical models albeit their weaknesses.

To date the most common ANN models used in the field of medicine includes the Multilayer Perceptron and the Radial Basis Function both of which can be calibrated to suit feed-forward, back propagation or recurrent neural network architecture to support clinical and non clinical decision making. According to

Sloane (1997) the first successful swarm intelligence to have been applied to the field of healthcare was the Ant Colony Optimisation (ACO). This concept was introduced by Dorigo et al (2006) to help solve discrete optimization problems in the late 1980s. The concept draws inspiration from the social behavior of ant colonies where a group of almost blind ants can jointly identify their shortest route between their food and their nest without visual information. Some of the common areas in the field of healthcare where the ACO has been applied include staff, patient and transportation scheduling, healthcare supply chain management, content based image retrieval, health insurance services etc (Collen & Ball, 2015)

In 1995, a renowned engineer Russell Eberhart and James Kennedy (a social psychologist) advanced the frontiers of swarm intelligence models when they successfully developed and applied a second concept of swarm intelligence known as the Particle Swarm Optimization (PSO).

Originally, the PSO was intended to help solve nonlinear, continuous optimization problems, but its application has expanded with time to many practical, real life problems. Not only is PSO used to track dynamic systems, evolve weights and structure of neural networks but is a major tool to analyse human tremor, register 3D-to-3D biomedical image, control reactive power and voltage and even learning to play games and music composition (Ryan et al., 2012). In the field of medicine, PSO is also a leading technique in solving multi-objective facility location-allocation problem for providing healthcare services, disease (cancer and myocardial) diagnosis and prognosis, predicting the length of stay of patients, classification of medical dataset etc Another crowd wisdom technique to have evolved and used in the field of healthcare is prediction markets.

Despite the major advantages and medical usefulness of the different crowd wisdom techniques examined earlier (prediction market, swarm intelligence, Delphi etc), their robustness is attenuated by many limitations identified over time with use. For example Lin et al (Lin et al., 2015) explains that the Delphi method is limited by the fact that it does not cope well with widely differing opinions (in this case medical) or large changes in opinions (paradigm shifts). Sometimes the facilitator's view may dominate in the analysis. Beyond

the large amount of time needed to engage in effective use of the method, it is also the contention of Farmer (Ryan et al., 2012) that differing opinions may not be sufficiently investigated. Moreover the success of the process and the quality of the outcomes is largely depended on high quality and highly motivated participants. More so known and evolving endemic limitations of swarm intelligence techniques and artificial neural networks disables them from being full proof in any medical decision making.

III. RESULTS AND DISCUSSION

Ethical and Moral Issues in Clinical Application of Crowd Wisdom Technique

However, like all medical decisions, the context specific applications of any decision that involves clinical judgment (such as the collection of the views of expert clinicians) strictly for medical decision making may be driven by legal, moral and ethical as opposed to medical reasons. This is acknowledged by Tiffin et al (2005) when they note that physicians, nurses, and other clinicians readily acknowledge being troubled by encounters with patients who trigger moral and ethical judgments with legal consequences. For decades social scientists have noted that ethical and moral judgment of patients is pervasive, occurring not only in egregious and criminal cases but also in everyday situations in which appraisals of patients' social worth and culpability are routine (Harrell Jr, 2015).

Therefore, it is imperative for clinicians and policy makers to understand how the conflict between legal, medical, and moral implications of clinical judgment shape the evolution and overall acceptance of the concept of clinical judgment in routine medical practice due to potential abuse and subsequent negligence by clinicians. In the extant literature the interplay between legal, moral and ethical considerations in clinical judgment in particular and medical practice in general has been debated with conflicting outcomes and interpretations. According to Pence et al (2004) at the heart of the legal, ethical and moral consideration lie the question of whether or not the practice of a clinician falls or has the propensity to fall below the required standard of care. Under common law, the minimal acceptable standard of care is measured against responsible medical practice, and usually not against guidelines (Milunsky & Milunsky, 2015). In Common

law, therefore, it is expert medical evidence that primarily assists the court in determining what the standard of care should be, and until now clinical guidelines have played a subsidiary role.

Eckles et al (2005) expound further on this notion when he argues that in the specific case of the moral, ethical, and legal application of clinician judgement that may result in negligence, the standard of care is judged by the *Bolam v Friern Hospital Management Committee* (1957) test with a subtle gloss added in the judgment in the *Bolitho case*. In *Bolitho v. City and Hackney Health Authority* (1996) the court declared that it was not bound to find for a defendant simply because he leads evidence from a body of experts who genuinely believe that the defendant's practice conformed to sound medical practice. The court will require further evidence that the practice proclaimed has a logical basis, and that the defendant practitioner has weighed up the benefits and risks. In other words, after *Bolitho* the defendant clinician in the case of perceived negligence in clinical judgement would have to justify his stance in addition to having this endorsed by similar responsible practitioners (Kitzinger & Kitzinger, 2015).

Also in *Lowry v. Henry Mayo Newhall Memorial Hospital* (1986) the claimants argued that the treating physician had arbitrarily deviated from the American Heart Association's guidelines for advanced cardiac life support by administering atropine rather than epinephrine. The defendant physician argued that guidelines were not mandatory and therefore could be overridden by clinical judgment in an individual case. The Appeal Court affirmed the judgment in favour of the defendant and did not see guidelines as being more persuasive than the facts of the case itself. Thus adherence to guidelines may not exonerate the defendant but sound clinical judgment may (Rosenfeld, 2004).

Although clinical guidelines may be acknowledged as relevant, the courts in the USA will take into account other sources of information in determining the standard of care, which would include factors such as the hospital's own procedures and policies, and expert evidence. The claimant in *Helling v. Carey* (1974), appealing from a judgment in favour of an ophthalmologist, argued that the customary standard of care presented was inadequate and therefore

unreasonable. In finding for the claimant, the Washington Supreme Court refused to be bound by widely endorsed clinical guidelines that formed the basis of the standard proclaimed by the defendant. Thus there is no absolute judicial deference to compliance with clinical guidelines (Farmer, 2004).

This means that the process of exercising clinical judgments in individual clinical decision making as well as the formulation of medical guidelines themselves must be based upon the principle of 'reasonableness and accountability' expected of a public body or officer. Thus the decisions must be publicly accessible, that the rationale must rest on evidence, that there is a mechanism for appealing against decisions, and that there is regulation of the process (Farmer, 2004).

IV. CONCLUSIONS AND FUTURE RESEARCH DIRECTION

In this study we have reviewed the evolution of the crowd wisdom theory and their application to the field of healthcare. Without doubt the work of Galton and his companions has inspired new initiatives in harnessing expert opinion to complement organisational and individual decision making. The benefits of the crowd wisdom in the healthcare sector based on the review have been enormous yet there remain many unsettled questions regarding its efficacy and efficiency in both clinical and non-clinical decision making. There are still outstanding questions regarding the moral, ethical, legal and other factors that must be addressed and contextualized. Further issues relating to judgmental bias of expert opinion that restricts its generalization are still scanty. Identifying a clear path to reduce perturbation of expert opinion from truth and incorporating appropriate evidence based techniques to standardize its use remains an adventure in transition. Future research is needed in this direction

V. REFERENCES

- [1]. Batchelor, Roy, and Pami Dua ((2015). Forecaster Diversity and the Benefits of Combining Forecasts. *Management Science*, 41, 68-75.
- [2]. Biglarian, A, Hajizadeh, E, Kazemnejad, A, & Zali, MR. (2011). Application of artificial neural network in predicting the survival rate of gastric

- cancer patients. *Iranian journal of public health*, 40(2), 80.
- [3]. Burke, Harry B, Goodman, Philip H, Rosen, David B, Henson, Donald E, Weinstein, John N, Harrell, Frank E, . . . Bostwick, David G. (1997). Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer*, 79(4), 857-862.
- [4]. Chen, Long-jian, Lian, Guo-ping, & Han, Lu-jia. (2007). Prediction of human skin permeability using artificial neural network (ANN) modeling. *Acta Pharmacologica Sinica*, 28(4), 591.
- [5]. Harrell Jr, Frank E. (2015). *Introduction Regression Modeling Strategies* (pp. 1-11): Springer.
- [6]. Hastie, Reid, & Kameda, Tatsuya. (2005). The robust beauty of majority rules in group decisions. *Psychological review*, 112(2), 494.
- [7]. Kitzinger, Celia, & Kitzinger, Jenny. (2015). Withdrawing artificial nutrition and hydration from minimally conscious and vegetative patients: family perspectives. *Journal of Medical Ethics*, 41(2), 157-160.
- [8]. Knight, Brian (2014). *Are Policy Platforms Capitalized into Equity Prices? Evidence from the Bush/Gore 2000 Presidential Election*. NBER Working Paper No. 10333
- [9]. Kurtzberg, Jerome M, & Levanoni, Menachem. (2002). Data mining techniques for enhancing medical evaluation: Google Patents.
- [10]. Lausch, Angela, Schmidt, Andreas, & Tischendorf, Lutz. (2015). Data mining and linked open data—New perspectives for data analysis in environmental research. *Ecological Modelling*, 295, 5-17.
- [11]. Mackay M (2011). Practical experience with bed occupancy management and planning systems: an Australian view. *Health Care Manag Sci* 2011 Feb;4(1):47-56.
- [12]. Mackay M, Lee M (2015) Choice of models for the analysis and forecasting of hospital beds. *Health Care Manag Sci* 2015 Aug;8(3):221-230.
- [13]. Milunsky, Aubrey, & Milunsky, Jeff M. (2015). *Genetic disorders and the fetus: diagnosis, prevention, and treatment*: John Wiley & Sons.
- [14]. Pence, Gregory E. (2004). *Classic Cases in Medical Ethics: Accounts of Cases that Have Shaped Medical Ethics, with Philosophical, Legal, and Historical Bacgrounds*.
- [15]. Rajakovich, David, & Vladimirov, Vladimir. (2009). Prediction markets as a medical forecasting tool: Demand for hospital services. *Journal of Prediction Markets*, 3(2), 78-106.
- [16]. Schwartz G (2008) Estimating the dimension of a model. *Ann Stat*. 2008, 6 (2): 461-464. 10.1214/aos/1176344136.
- [17]. Shi, Hon-Yi, Lee, King-Teh, Lee, Hao-Hsien, Ho, Wen-Hsien, Sun, Ding-Ping, Wang, Jhi-Joung, & Chiu, Chong-Chi. (2012). Comparison of artificial neural network and logistic regression models for predicting in-hospital mortality after primary liver cancer surgery.
- [18]. Shi, Hon-Yi, Tsai, Jinn-Tsong, Chen, Yao-Mei, Culbertson, Richard, Chang, Hong-Tai, & Hou, Ming-Feng. (2012). Predicting two-year quality of life after breast cancer surgery using artificial neural network and linear regression models. *Breast cancer research and treatment*, 135(1), 221-229.
- [19]. Sloane, Neil JA. (1997). *Delivery of medical services using electronic data communications: Google Patents*.
- [20]. Spelt, Lidewij, Nilsson, Jessica, Andersson, Roland, & Andersson, Bodil. (2013). Artificial neural networks—A method for prediction of survival following liver resection for colorectal cancer metastases. *European Journal of Surgical Oncology (EJSO)*, 39(6), 648-654.
- [21]. Surowiecki, James. (2005). *The wisdom of crowds: Anchor*.
- [22]. Tiffin, Nicki, Kelso, Janet F, Powell, Alan R, Pan, Hong, Bajic, Vladimir B, & Hide, Winston A. (2005). Integration of text-and data-mining using ontologies successfully selects disease gene candidates. *Nucleic acids research*, 33(5), 1544-1552.
- [23]. Wolfers, Justin (2008). *Polls Glance Backward, Markets Peer Ahead*. *The Wall Street Journal*, July 8, 2008.
- [24]. Wong H, Wu RC, Tomlinson G, Caesar M, Abrams H, Carter MW, et al (2009) How much do operational processes affect hospital inpatient discharge rates? *J Public Health (Oxf)* 2009 Dec;31(4):546-553