**Medicine**

So far we have documented serious limitations in forecasting in social sciences, concluding of the low accuracy of statistical models but even then that their accuracy is superior, on average, that of judgment. Are medical predictions any different, given the huge costs involved and the much greater harm of benefits that can be caused by accurate or wrong predictions? Moreover, given the considerable advances in medical technology and the huge spending on medical research, including that of more effective medical drugs are there any concrete, objective benefits available?

In medicine there are three types of forecasts required. The first relates for the doctor to make the correct diagnosis. The second, to match such a diagnosis with the correct treatment to be decided by the doctor by matching it with the recommendations of the most appropriate therapy based on published research findings that are often presented in the forms of guidelines written by professional committees of doctors. The big question is how reliable are such findings as new research seem to invalidate old ones. Finally, there is also what is called *preventive* medicine that recommends various procedures and tests to identify potential, future problems before they become serious and take actions to deal with them at present. This section looks at each of these three types of medical predictions and discusses their successes and failures. Needless to reaffirm that their outcome is uncertain as is the case with all forms of future predictions.

***Diagnostic forecasting:*** The diagnosis of disease has progressed a great deal with the wide availability of laboratory tests (e.g. blood and urine), equipment like x-rays, ultrasound and MRI machines and PET and CT scanners but still research shows significant concerns. In a 2013 article, Ashley et al. state: “*A total of 118 physicians with broad geographical representation within the United States correctly diagnosed 55.3% of easier and 5.8% of more difficult cases (P < .001)* (making an overall average of a 31% success rate)*. Despite a large difference in diagnostic accuracy between easier and more difficult cases, the difference in confidence was relatively small (7.2 vs 6.4 out of 10, for easier and more difficult cases, respectively) (P < .001) and likely clinically insignificant. Overall, diagnostic calibration was worse for more difficult cases (P < .001) and characterized by overconfidence in accuracy”* (p. 1952). The conclusion of this study is not different than what is known in other areas of forecasting a dismal diagnostic record (although not directly comparable, this 31% overall rate of physician accuracy is lower than the results from a survey (to be used as a benchmark) reported in ***Isabel Healthcare*** Blog (Reference) that showed that 41% of patients ‘own diagnoses were confirmed by their physicians).

A major finding of the Ashley’s et al study was how little the physicians level of confidence changed from the easy to hard ones (7.2 out of 10 for the easy ones and 6.4 out of 10 for the hard ones). This means that with an accuracy rate of only 5.8%, the physicians were still 64% confident that they were right! A low diagnostic accuracy of 5.8% could be tolerated if the physician say was only 10% confident of being right as it would be more likely to order more tests or ask for a second opinion from another doctor. But a 64% confidence would probably exclude such actions and would probably proceed with the wrong treatment. [Berner](http://www.ncbi.nlm.nih.gov/pubmed/?term=Berner%20ES%5BAuthor%5D&cauthor=true&cauthor_uid=18440350) and [Graber (2008)](http://www.ncbi.nlm.nih.gov/pubmed/?term=Graber%20ML%5BAuthor%5D&cauthor=true&cauthor_uid=18440350) further discuss how overconfidence results in diagnostic errors and what needs to be done to reduce them. It seems that the common problem of overconfidence in forecasting in general is also present in medical diagnosis with its detrimental negative consequences (see below for suggestions of elevating such negative consequences).

There are other publications studying diagnostic errors. [Graber,](http://www.ncbi.nlm.nih.gov/pubmed/?term=Graber%20ML%5BAuthor%5D&cauthor=true&cauthor_uid=16009864) [Franklin and](http://www.ncbi.nlm.nih.gov/pubmed/?term=Franklin%20N%5BAuthor%5D&cauthor=true&cauthor_uid=16009864) [Gordon (](http://www.ncbi.nlm.nih.gov/pubmed/?term=Gordon%20R%5BAuthor%5D&cauthor=true&cauthor_uid=16009864)2005) conclude: *“Diagnostic error is commonly multifactorial in origin, typically involving both system-related and cognitive factors”* (p. 1493) while Tehrani et al. (2013) reports that *“Among malpractice claims, diagnostic errors appear to be the most common, most costly and most dangerous of medical mistakes”* and continue by suggesting that “*healthcare stakeholders should consider diagnostic safety a critical health policy issue”* (p. 672). In another study about diagnostic errors in radiology Lee et al. (2013) state: *“The retrospective error rate among radiologic examinations is approximately 30%, with real-time errors in daily radiology practice averaging 3–5%. Nearly 75% of all medical malpractice claims against radiologists are related to diagnostic errors”* (p. 611).

Fink, Lipatov and Konitzer (2009) claim that the prediction and reliability of diagnoses follow a power law with a “Black Swan” element in the majority of consultations. If this is true it adds a whole other dimension to the diagnostic process, increasing uncertainty considerably and requiring a novel approach to the diagnosis of low incidence disease. The critical importance of diagnosis and the need to improve its accuracy has been recognized in recent years. Newman-Toker and Pronovost (2009) their article ***Diagnostic Errors—The Next Frontier for Patient Safety***, as the title describes, present an overview of the importance and magnitude of diagnostic errors and provide suggestions for improvements. [Along similar lines, Singh](http://qualitysafety.bmj.com/search?author1=H+Singh&sortspec=date&submit=Submit), [Petersen](http://qualitysafety.bmj.com/search?author1=L+A+Petersen&sortspec=date&submit=Submit) and  [Thomas](http://qualitysafety.bmj.com/search?author1=E+J+Thomas&sortspec=date&submit=Submit) (2005) mention that diagnostic errors have both a cognitive and a system components and are not well understood, making their prevention unlikely. Their suggestion is a situational awareness conceptual model in order to better study and be able to remedy the shortcoming of the current diagnostic process.

There is another potential problem: too much diagnosis. In an article in New Yorker, Gawanda (2009) compares the medical costs in two cities McAllen and El Paso. He writes that “*between 2001 and 2005, critically ill Medicare patients received almost fifty per cent more specialist visits in McAllen than in El Paso, and were two-thirds more likely to see ten or more specialists in a six-month period. In 2005 and 2006, patients in McAllen received twenty per cent more abdominal ultrasounds, thirty per cent more bone-density studies, sixty per cent more stress tests with echocardiography, two hundred per cent more nerve-conduction studies to diagnose carpal-tunnel syndrome, and five hundred and fifty per cent more urine-flow studies to diagnose prostate troubles. They received one-fifth to two-thirds more gallbladder operations, knee replacements, breast biopsies, and bladder scopes. They also received two to three times as many pacemakers, implantable defibrillators, cardiac-bypass operations, carotid endarterectomies, and coronary-artery stents. And Medicare paid for five times as many home-nurse visits. The primary cause of McAllen’s extreme costs was, very simply, the across-the-board overuse of medicine”*.

(<http://www.newyorker.com/magazine/2009/06/01/the-cost-conundrum>)

In his NYT article (Welch, et al., 2007) doctor Gilbert Welch and his coauthors explain how test results make people sick and why visiting a doctor can be hazardous to your health. Their advice is to avoid all preventive medical care.

There are several proposals for improving diagnosis, reduce costs and avoid serious mistakes and patients’ harm. The Internet site ***Society to Improve Diagnosis is Medicine*** offers a complete coverage of all aspects of medical diagnosis also offering suggestions for improvement. (<http://www.improvediagnosis.org/?page=Facts>) Another Internet site aiming at improving the process of diagnosis is **Isabel**. These Internet tools can be used as aids and/or supplement to doctors’ diagnosis by having the patient to enter their details before the consultation and become active participants during the consultation session. (<http://www.isabelhealthcare.com/home/default>) An alternative could be computer programs taking into account all symptoms and coming up with an objective diagnosis that could be compared and supplement that of the doctors.

***How reliable are the findings of medical research?*** The following quote is from ***Time*** Magazine’s March 21, 2012 issue: *“Many people take a daily aspirin to reduce their risk of heart attack, but now fresh evidence suggests that the over-the-counter pain reliever may be a powerful tool in cancer prevention as well. In three new studies published in the****Lancet****, researchers from the University of Oxford say a daily dose of aspirin can reduce people’s risk of developing a variety of cancers and also lower the chance of their cancer spreading”* (Alice Park, 2012). Given such favorable reports, millions of people around the world take daily low-dose aspirin to benefit from this wonder drug that, in addition, costs just a few cents.

But practically concurrently with the ***Time***’s article appear another one in the ***Archives of Internal Medicine*** (Seshasai, 2012) that concluded: “*Despite important reductions in nonfatal MI, aspirin prophylaxis in people without prior CVD does not lead to reductions in either cardiovascular death or cancer mortality. Because the benefits are further offset by clinically important bleeding events, routine use of aspirin for primary prevention is not warranted and treatment decisions need to be considered on a case-by-case basis” (p.209).* A more recentarticle (Hira, 2015) concludes that the risks associated with regular aspirin use outweigh the benefits. So what should people do? The recommendation is to ask their doctor. But does he or she can know any better, given the conflicting evidence and reliability of medical research findings (see below)?

Similar reversals have been observed with coffee too. Originally, research findings were negative, concluding to an association of coffee drinking with coronary heart disease that was increasing with a higher the coffee consumption (LaCroix et al., 1986) as well as an increase in the risk of death (Freedman et al., 2012). Later findings, however, showed benefits that ranged from a longer life expectancy (Butt and Sultan, 2011) to reducing ovarian cancer (Turoger, et al., 2008). So what should people do, will the most recent findings hold or will be reversed again by new ones?

Going back and forth in aspirins and coffee does may not involve life threatening situations but *bone marrow transplant* definite does. In addition to involving an extremely painful procedure, it is also downright expensive, with costs ranging upwards to $150,000 and reaching the million marks in complicated cases (Brawley, 2011). The idea behind the marrow transplant was to harvest and store it in order to avoid the toxic effects of chemotherapy and then re-introduce it was completed. Well, randomized trials found that the procedure did not improve survival and it was, therefore, abandoned (Brawley, 2011, p.35). Unfortunately, the marrow transplant disaster is not unique in medical research.

Ioannidis has published widely on the deficiencies of medical research (2005, 2005, 2010). In his PLoS Medicine article he states *“There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims”* (p. 696). In his article in ***JAMA*** (2005), he concludes *“Contradiction and initially stronger effects are not unusual in highly cited research of clinical intervention and their outcomes”* (p. 218). In a 2010 article in the ***Atlantic***, featuring Ioannidis, David Freedman quotes him saying “*that as much as 90 percent of the published medical information that doctors rely on is flawed and that he worries that the field of medical research is so pervasively flawed, and so riddled with conflicts of interest, that it might be chronically resistant to change—or even to publicly admitting that there’s a problem”.* This and the previous quotes speak for themselves, requiring no additional comments. *(*[*http://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/*](http://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/)*)*

From a forecasting point of view the critical question is how some research finding can be utilized to base therapy when a future, new one could reverse its recommendations? Is there something fundamentally wrong with the practice of medicine that requires a fundamental rethinking of how it is practice? Should the medical community admit that there is a serious problem facing the profession? In addition, is it possible to ensure that conflicts of interests will not influence the diagnostic process and the recommended therapy as the case of Caesarian Section (CS) proves?

There has been a considerable increase in the number of CS during the last few decades in some countries like Greece and Brazil exceeding the three fourths of all births and becoming even higher on women with private insurance. A recent report of the World Health Organization (WHO) concludes (Reference, 2015):

*Based on the available data, and using internationally accepted methods to assess the evidence with the most appropriate analytical techniques, WHO (The World Health Organization) concludes:*

*1. Caesarean sections are effective in saving maternal and infant lives, but only when they are required for medically indicated reasons.*

*2. At population level, caesarean section rates higher than 10% are not associated with reductions in maternal and newborn mortality rates.*

*3. Caesarean sections can cause significant and sometimes permanent complications, disability or death particularly in settings that lack the facilities and/or capacity to properly conduct safe surgery and treat surgical complications. Caesarean sections should ideally only be undertaken when medically necessary.*

Given the clarity of recommendations/evidence it is strange the high rates of actual caesareans simply because of convenience reasons and because the increase the revenues of doctors while ignoring the potential damages to mothers and children.

***Preventive medicine:*** According to the independent Cochrane Foundation *“General health checks involve multiple tests in a person who does not feel ill with the purpose of finding disease early, preventing disease from developing, or providing reassurance. . . . To many people, health checks intuitively make sense, but experience from*[*screening*](http://summaries.cochrane.org/lexicon/9#screening)*programs for individual diseases have shown that the benefits may be smaller than expected and the harms greater. A possible harm from health checks is the diagnosis and treatment of conditions that were not destined to cause symptoms or death”* (p. 2). For instance yearly checkup examinations started in the early 1920s and have continued since then, although many studies going back to the 1960s have shown no benefits from them. For instance Krogsboll et al. (2012) concluded: “General health checks did not reduce morbidity or mortality, neither overall nor for cardiovascular or cancer causes, although the number of new diagnoses was increased” (p*. 2).* Yet, despite evidence against routine annual examinations, many family physicians recommend them (Mehrotra, Zaslavsky and Ayanian, 2007) exploiting the *“illusion of reassurance”* that a preventive test will catch health problems early, reducing disease and increase life expectancy. But this has not been the case

Another frequent recommended preventive test is annual mammography for all women older than 40 (it is estimated that 85% of women in the USA each year are screened for breast cancer). Lately, the starting age has been raised to 50 and the interval of the screening to two years instead of annually. But there is still an intense argument for the value of any screening for women at any age with those opposing it saying that there is strong evidence that the potential harm from screening is considerably greater than the benefits. Gøtzsche in his book (Gotzsche P.C., 2012) “***Mammography Screening: Truth, Lies and Controversy***” states:

*“If we wish to reduce the incidence of breast cancer, there is nothing as effective as avoiding getting mammograms. It reduces the risk of getting breast cancer by one-third.” and later he continues “We have had devastating epidemics of hysterectomies and tonsillectomies and we currently have an epidemic of Caesarean sections, although these operations may lead to substantial harm. We have also epidemics of mastectomies and prostatectomies because people seem to be more worried about dying than they are interested in living.”*(p.349)

Prostate cancer tests for men are also common as it is estimated that close to 52% of men are tested annually at a cost exceeding half a billion for Medicare patients alone in the USA. Richard Ablin, the inventor of the PSA test used to diagnose prostate cancer, in his book (Ablin, 2014) The Great Prostate Hoax states:

***“****The ability of the PSA test to identify men with prostate cancer is slightly better than that of flipping a coin. And its continued use as a routine screening tool is nothing short of a national health disaster”* (p. 6). And later he continues *“Among the 1,000 men who had the PSA test, 20 of them would have radical prostatectomies for cancers that never would have caused symptoms. And five of those men would have lifelong complications, including impotence and incontinence”*(p. 45).

Later in his book, Albin states that if the probability of men dying from prostate cancer is 3% a year, that of surviving a diagnosis of prostate cancer is 97% whether he receives treatment or not. Then why should a man be screened if the harm of false diagnosis is 30 to 100 times the estimated benefit, resulting in less than 0.1% reduction in prostate mortality over ten years? His conclusion is that the screening is done because of financial interests as it increases the number of additional tests and prostatectomies. He asserts that *“Without radical prostatectomies, more than half of all the urology practice in the United States would go belly-up”*(p. 42).

Preventive testing is another medical area requiring fundamental rethinking with the advice of epidemiologist McPherson being that *“reducing incidence* (diagnosis of cancers) *must be the primary goal, with reducing mortality an important but secondary end point*” (p. 233-5). (SEE P. 361, Gotzsche). The question is if vested interests from both doctors and pharmaceutical firms would allow any changes in established practices?