#10 Physicians’ Use of Probabilistic Information in a Real Life Clinical Setting

A Critical Appraisal

Paper found in a list cited on the webpage:
http://www.staff.ncl.ac.uk/mike.cox/PsyFin/assbiblio.pdf

Reference:


University of Newcastle-Upon-Tyne
Is the study question relevant?

The study examines qualified physicians’ use of base-rate information and predictive values in diagnosing possible pneumonia patients. It is based in a real-life setting, investigating the diagnosis and resulting treatment of a patient. The study examines flawed diagnostic techniques to improve this process. The results identify a cognitive flaw in decision making in relation to psychology whilst real-life rehabilitation of sick patients is addressed. The outcome suggests physicians overestimate probability of pneumonia, highlighting a need for reassessment of diagnostic techniques and cognitive processes used when identifying serious illnesses.

A specific hypothesis is not identified and the paper is referred to as a “report”; a strange classification for a paper included in such an esteemed journal. This lack of focus is problematic; the findings are not related back to theories tailored to the study, only to previous literature. The inclusion of a hypothesis clearly stating the aims of the research may give the paper better structure and would allow experimenters to be precise in their measures and explicitly analyse findings in relation to predicted outcomes.

Does the study add anything new?

Critically, the study improves on previous laboratory experiments by addressing probability bias in a real-life setting. Past research has investigated the effect amongst unqualified participants who have no commitment to ensure their prediction is accurate; Christensen-Szalanski & Bushyhead’s (1981) study employs experts who have an obligation to provide accurate diagnosis in order to begin appropriate treatment. Therefore, the findings of this study are more ecologically valid and reliable than antecedent experiments with regards to using faithful participants dedicating a greater amount of time and energy to the decision-making process. No previous literature reports on the validity of physicians’ probability estimates in comparison to the degree of calibration (flawless diagnosis).

The authors discredit physicians’ diagnostic abilities by exposing a critical flaw in the external validity of their probabilistic determinations. The results indicate a significant gap between predictions and perfect calibration, indicating overestimation of probability of pneumonia. The study additionally examined the ability to estimate probability of pneumonia by considering absent symptoms, a measure which greatly
increases the accuracy of diagnosis. Previous research has highlighted a cognitive inadequacy in the processing of absent cues,\(^2\),\(^8\),\(^10\) yet this paper incorporates the claim to explain the exact causes of the overestimation bias.

In situations for which a correct decision is not based on expertise and knowledge, participants are inclined to reject the base-rate statistics in favour of intuition.\(^1\),\(^4\),\(^10\),\(^11\) Even in experiments using similar qualified participants, findings have suggested that experts cannot utilise base-rate information correctly.\(^2\) However, the present study indicates that physicians incorporate base-rate information underpinned by experience, allowing employment of a calculative method of prediction due to increased salience of these statistics.

**Was the study design appropriate for the research question?**

The multiple-choice parameter common to previous laboratory studies was removed hence changing the payoff matrix, increasing accuracy of predictions. Instead, participants used expertise to combine historical and physical medical attributes of the patient to decide diagnosis. This method seems more appropriate for the study as it removes cognitive parameters. Conversely, the removal of this measure encourages a lack of control; along with the natural confounding factors present in all real-life settings, this poses a threat to validity as there are few constraints on the variables affecting cognitive processes.

Had each physician examined each patient their decisions could be compared, allowing causal identification of anomalies; reigning in control and introducing inter-rater reliability. This would produce a sample closely resembling real life as physicians are likely to see a number of patients daily: Tooby & Cosmides\(^9\),\(^3\) suggest that natural frequencies increase likelihood of correct Bayesian reasoning (76-92%). A study replication with this measure enforced may observe a rise in diagnostic accuracy and decreased overestimation.

**Did the study methods address the potential sources of bias?**

Physicians used a checklist to collect medical history, installing information retrieval standardisation. Conversely, this measure heightens attention to potentially relevant features, discrediting the validity and authenticity of the results in accordance with everyday cognitive processes that occur in decision making.

The cross-comparison of physicians’ diagnosis with X-ray examination is reassuringly unbiased as radiologists had no prior knowledge of patients’ medical history, physical symptoms (except a minor cough)
or diagnosis. The authors admit that physicians’ immediate perception of the patient being in poor health as a result of the patients’ decision to visit a clinic may have skewed the results and may in fact be the main cause for the findings that predictive value is underestimated yet probability of pneumonia is overestimated. 

A replication of the study where physicians are presented with a number of random individuals and their corresponding ailments (if there are any) is suggested. The results of the two studies should be compared to redeem the ecological validity of the existing study whilst reducing the preliminary bias in the follow up.

**Are the data justified?**

The authors initially appear to make overconfident claims, later admitting that a single outlier severely skews the results of the symptom absent condition, and find that the regression equations for each condition are almost identical once it is eliminated. Even without removing this anomaly, a t-test indicates no significant difference between conditions. A revision of analysis is recommended: perhaps a log scale would be more appropriate to home in on the data.

Conversely, anomalous data is still valid data; the exclusion of outliers would be misleading. The authors have therefore mentioned a flaw in their work which is not in fact a fault.

**Discussion**

The opposing finding that physicians incorporate the use of base-rate information underpinned by experience highlights a need for further investigation and defines a differential decision-making process between professionals and uninformed individuals. This innovative discovery credits the experiment with originality by testing the use of base-rate information amongst experts. However, this same factor undermines the ability to generalise the findings to a wider population, isolating it from the universal study of decision-making.

Upon repetition of enforcement of probabilistic beliefs, most individuals are incorrect 15-20% of the time (5, 7) yet many expect to correctly calibrate 98% of their predictive values and are significantly overconfident in their estimates (5, 6). An analysis of the data representing physicians’ assignment of pneumonia probability and personal value rating of each diagnosis outcome would give insight into their diagnosis certainty
compared to accuracy, indicating any signs of overconfidence. It can be predicted that physicians will be overconfident in their diagnoses as they are trained experts used to finding that there is at least some minor explanation for patients' ailments, be it pneumonia or another cause; this is suggested by the overestimation effect observed in the study. Again, this brings in base-rate information: physicians will probably detect a present illness due to the usual positive outcomes of examinations they have experienced. The results of this analysis will highlight any overconfidence of physicians in their predictions and indicate whether there perhaps should not be such strong regard as some individuals place on diagnoses.

The proposed reason for such errors is the notion that physicians do not receive much feedback. Those who receive regular comment on their judgements exhibit little or no overconfidence – physicians are not likely to experience this as they are not always explicitly informed of the accuracy of their decision. Meteorologists, for example, receive constant feedback (from the weather and the media) and show perfect calibration.

A study relating this overconfidence effect to finance could be conducted analysing drivers’ attitudes to car insurance. As the locus of control is internal in a driving situation, the decision not to take out fully compensated insurance would indicate confidence in driving ability. Purely examining claim rates of those with alternative policies to fully compensated would indicate the degree of overconfidence. In a controlled experiment, stated confidence in driving ability of those without full insurance could be compared to number of unclaimed crashes and claims. Anything other than a negative correlation between confidence and crashes would signify overconfidence. This information could be utilised in insurance to provide information on the demographic of typically overconfident individuals. From this, companies can individually tailor policies and target this group to warn them about flawed judgement and potential dangers.

It has been claimed that overconfidence is one the “most prevalent and potentially catastrophic problem in judgement or decision making”, hence explaining why physicians seem likely to overestimate the probability of present pneumonia, erring on the side of caution in such crucial situations. It is for this reason that further studies on overconfidence, are inherent in ensuring the safety and wellbeing of a great manner of events.
References


