IMPROVING FORECASTING AND PREDICTION IN CLINICAL PSYCHOLOGY: WHAT CAN BE DONE TO IMPROVE ACCURACY?
1. INTRODUCTION: PREDICTION & ACCURACY ............ 3
   IN CLINICAL PSYCHOLOGY
2. RESEARCH IN FORECASTING ON IMPROVING ........ 4
   ACCURACY, AND HOW IMPROVED ACCURACY
   CAN BE JUDGED
3. ACTUARIAL PREDICTION: METHODS, ...............6
   ADVANTAGES, AND LIMITATIONS
4. EXPERT SYSTEMS IN CLINICAL PSYCHOLOGY: ........8
   RESEARCH AND FUTURE STEPS
5. COMBINING ACTUARIAL AND JUDGEMENTAL ..........11
   METHODS: EXPERT ADJUSTMENTS
6. STRUCTURED JUDGEMENT ...............................12
7. CONCLUSION ................................. 13
8. REFERENCES ................................. 15
INTRODUCTION: PREDICTION & ACCURACY IN CLINICAL PSYCHOLOGY

Research (Meehl, 1954; 1986) has consistently demonstrated that a number of problems exist with predictions made by clinical practitioners such as psychologists and doctors about an outcome, be it physical or psychological illness, educational attainment or future criminal behaviour. In this sense prediction refers to how data (be it quantitative or qualitative) is processed or aggregated (Westen & Weinberger, 2004) to form a decision about a state of affairs in the present and future. In this sense, prediction in clinical psychology is very much similar to prediction in other areas of forecasting, as there is widespread uncertainty about not only the future, but the present as well. Take the example of depression; the DSM-IV definition states that the symptomology associated with the disorder must persist for longer than two months (DSM-IV, 2000), and must have 5 of the symptoms outlined. Ancillary to this is how depression is regularly comorbid with anxiety, and can be related to self harming and suicidal cognitions and behaviours; the DSM-IV notes the latter to be one of the key symptoms of depression, although the relationship between the two is not perfect. At each of these stages the clinical psychologist must make a prediction: they must judge from the data presented the likelihood the patient fulfils the criteria of depression, whether they have comorbid disorders, and the likelihood they commit specific harmful behaviours. When one thinks of the sort of things a clinical psychologist does, this example perhaps helps to furnish how pervasive prediction is within clinical psychology, a sentiment shared by Faust (2004), who makes a distinction between implicit (such as treatment efficacy) and explicit (risk of violent/suicidal behaviour) predictions. Grove & Meehl (1996) note that even the actions a clinician performs which do not involve a prediction are predicated on a prediction.

The crucial finding of Meehl’s research, and nearly six decades of research following it, is that actuarial (or statistical) methods can reliably perform at the same standard or greater than a prediction made by a practitioner, and can do so with greater reliability. Moreover (Meehl, 1956; Dawes et al., 1989; Grove et al., 2000), the relative efficacy of the actuarial approach is greater when the practitioner has more information and particularly the opportunity to interview a patient. Generally, the actuarial or mechanical prediction is accurate by about 10% (Grove et al., 2000) more than a
clinical prediction. Whilst this may seem relatively small, it must be stressed that in the context of forecasting this is a considerable improvement. Research within forecasting has often lamented the difficulty in predicting future events (Makridakis & Taleb, 2009), and as such, an increment of this value should be appraised appropriately.

The evidence appears therefore to suggest that clinical prediction can be improved, and also there is a fair amount of inaccuracy in current clinical judgemental practices. This is problematic; practitioners, such as clinical psychologists, are required and valued to be adept at forecasting present and future patterns of human behaviour, both to ensure that a disorder is correctly identified, and that the proposed means of treatment are appropriate and effective. A mistaken forecast can lead to a psychological, pharmacological (particularly in the US where clinical psychologists have limited abilities to prescribe psychiatric medications) or medical treatment that may be ineffective or iatrogenic, or worse a patient committing a violent behaviour towards oneself or another. Moreover whilst the objects of prediction (such as patients or behaviours) are highly difficult to predict, clinical psychologists also have the burden of a high expectation of high accuracy by laypeople. Shanteau (1992) notes also that the general public place very high values on accuracy for experts such as psychologists, finance experts and psychiatrists, despite the highly volatile and unpredictable nature of the subjects of prediction.

As such, the findings of this field of research beg the question: what can be done to demonstrably improve clinical accuracy? This dissertation aims to answer this in a proverbial two-pronged assault. The first is to briefly review how forecasting research aims to improve prediction, and apply the recommendations it generates about judgemental forecasting to clinical settings, building on research where this exists. The second is to examine where current practice can be improved to maximise the accuracy of forecasts made by clinical psychologists.
RESEARCH IN FORECASTING ON IMPROVING ACCURACY, AND HOW IMPROVEMENT CAN BE JUDGED

Considering the forecasting literature and its applications to clinical literature, it is important to consider the broad nature of forecasting. Most importantly, judgemental methods have been and to some extent still are largely marginalised; classic (Hogarth & Makridakis, 1981; Makridakis, 1988) and contemporary (Makridakis & Taleb, 2009; Armstrong, 2006) research has placed caution upon the use of judgemental methods, citing their poor accuracy, and research has instead focused upon statistical forecasting. However, judgemental forecasting has become more prevalent; Lawrence and colleagues (2006) note that in business and finance, judgemental and actuarial means are widely used in tandem, in part because purely statistical techniques have been felt to be imperfect by non researchers. Other researchers in judgement and forecasting, such as Fischoff (1988) have noted that actuarial methods contain considerable elements of judgement anyway (a view strongly supported by Weston & Weinberger, 2004).

Research reviewing forecasting accuracy has provided several clear principles concerning how to improve forecasting accuracy; Armstrong (2006) notes that a conservative approach ought to be adopted when making a prediction: this involves spreading risk (so for example a wide range of empirically validated measures to maximise sensitivity), large amounts of data, prediction models based on previous research and the use of structure, all of which can be applied to clinical practice.

Other recommendations are less practical; one example of these is the use of feedback (Lawrence et al., 2006). This is difficult to apply to clinical prediction for several reasons. The most obvious is that there is little opportunity for feedback of diagnoses or treatment efficacy; many patients that are referred to a psychologist will remain in treatment for years. A psychologist will generally focus upon the disorder that is of distress to the patient, and they may fail to diagnose other disorders. Thus feedback items such as accuracy may be less salient to the psychologist than other experts.

Important to talk about when discussing improving clinical prediction is the concept of incremental validity. As will be demonstrated in the remainder of this paper, many of the possible applications of an improvement from forecasting or elsewhere require
the addition of something else; a test, the addition of structure to an interview or an actuarial equation. Where incremental validity is important is that these additions show incremental validity if they can explain the true state of affairs over and above the means currently used (Hunsley & Meyer, 2003), in this case being clinical judgement. To furnish this with an example, let us return to the example of depression. There are a number of tests produced for interview and psychological assessment such as the Beck Depression Inventory (Beck et al., 1961), Hamilton Depression Scale (Hamilton, 1960), Structured Clinical Interview for DSM-IV (First et al., 2004), HADS (Zigmond & Snaith, 1983) or the Geriatric Depression Scale (Yesavage et al., 1983). Would it be useful for a psychologist to carry all of these out? Of course not; irrespective of time and context, much of the additional information gathered would be redundant, and have little incremental validity. Of importance to this dissertation therefore, is that the suggestions outlined must clearly demonstrate incremental validity.

Before exploring these, it is important to note one important caveat: an obvious explanation is that there may be considerable variance in the quality of the expertise, and this variance may account for weaknesses in clinical prediction. More specifically, actuarial prediction may outperform clinical prediction, but the best clinicians may outperform the actuarial means. However, there appears to be little research to support this assertion. Research (APA, 1982; see Einhorn, 1972 for an example) has concluded that expertise and predictive accuracy are largely unrelated. Garb (1989) in a review article investigated the differences at different levels of expertise in assessment accuracy. Sections compared seasoned and inexperienced experts and graduate students, at no point finding any evidence for a difference in performance. If the learning and experience that presumably come with being a clinician are independent of clinical accuracy, then other means must be sought to improve accuracy. At any rate, as the educational expertise required of a clinical psychologist has decreased over time (Dawes, 1994), this does not appear to be a realistic solution anyway.
ACTUARIAL PREDICTION: METHODS, ADVANTAGES AND LIMITATIONS

The findings of Meehl’s work, and the continuation by other researchers explicitly points to the use of actuarial, or statistical decision rules in improving clinical prediction accuracy. Widely cited examples of these include the Goldberg Rule (Goldberg, 1965; Dawes et al., 1989) in determining whether a patient is psychotic or neurotic using the Minnesota Multiphasic Personality Inventory (MMPI). These are rules produced by clinical experts (Garb, 1994) that systematically weigh collected data to discern between multiple outcomes. To furnish this with an incredibly simple example, take the Beck Depression Inventory (Beck et al., 1961). This is a series of 23 questions that are intended to represent the criterion of depression. These questions are rated from 0-3, which attempts to tap into whether the patient is depressed or not.

Of course actuarial systems which are intended for use in clinical research are usually a lot more complex than this, taking data from a number of different sources, not limited to psychometric instruments such as the BDI, but also interview data, self/family reports or projective data such as drawing or Rorschach interpretations. The BDI has several cut-off points; a score >9 indicates mild depression, >20 moderate, >30 severe, each of which discerns a different outcome. An actuarial means will work in a similar fashion, although considerably more complex to account for multiple data sources, different weights applied to them (Wiggins, 1981; Dawes & Corrigan, 1974) and different instruments.

However, these themselves have been criticised for a number of reasons (Garb, 2000) including lack of statistical power, research that has utilized only a single data source, has not always demonstrated incremental validity (Hunsley & Meyer, 2003) and is sometimes weaker than expert clinical judgement (particularly in the context of brain damage). They have also been demonstrated to be poor at predicting future behaviours of critical importance to clinical psychology, such as future incidence of suicide. However, other research concerning violence for example (Elbogen, 2002; Webster et al., 1997) has found the opposite. Moreover research using them has generally not integrated psychometric & qualitative data, opting to generally use the former (Garb, 1994), meaning that the incremental validity of the statistical rule may not be entirely demonstrated in the research carried out. Generally however research
has demonstrated the efficacy of the actuarial prediction, and many of the criticism have been largely addressed (Grove & Meehl, 1996).

A further issue raised from both the forecasting and clinical literatures is the need to test out actuarial methods on new cases; forecasting research (in the form of the M competitions; e.g. Makridakis & Hibon, 2000) has found that highly sophisticated models (such as ARMARA or ARIMA) that fit previous data better than less complex method ended up being less able at forecasting future outcomes, although this is generally not applicable to the clinical actuarial research carried out.

Whilst these give relatively strong evidence for the use of actuarial rules, it is not unanimous. The most important aspect is that such decision rules appear to be maximally efficacious when using information to predict a more general pattern of a patient’s state of mind, but not specific outcomes. Equally however this cannot necessarily be found to be a weakness of actuarial methods in general; one of the main criticisms of actuarial rules seems to lie in areas where research or measures are limited or poorly validated. To this end some (Swets et al., 2000; Lilienfeld, 2004) have argued that actuarial rules that research has demonstrated to be well validated are used in conjunction with clinical judgement. Regardless, the evidence suggests that a transition towards actuarial means is supported by research evidence. In particular, it can be said that this ought to produce a considerable improvement in clinical prediction, given that less than a third (31%) of practitioners use actuarial methods in prediction (Vrieze & Grove, 2009), despite nearly 60 years of research endorsing them. This stands in contrast to predictions made by Goldberg (1970), who suggested that actuarial methods might be adopted by psychology whilst viewed with suspicion in areas such as medicine and finance. Vrieze & Grove’s study, combined with the frequency of research in financial forecasting using actuarial methods (i.e. Makridakis & Taleb, 2009) suggests the exact opposite has happened. This raises two issues. The first is the non usage of a scientifically validated means of prediction ought to suggest that movements towards its widespread adoption should considerably improve clinical prediction. The second is that there appears to be considerable resistance to using actuarial means. This means that in order to improve prediction, alternatives that place a focus upon more judgemental approaches may have to be considered on a pragmatic level to maximise adoption.
EXPERT SYSTEMS IN CLINICAL PSYCHOLOGY: RESEARCH AND FUTURE STEPS

One possible means of improving clinical predictions is to transform aspects of a clinical prediction into an actuarial prediction by way of an expert system. An expert system is a decision maker or aid, which using knowledge acquired from experts represents the decision making behaviour of the expert in the domain they hold expertise in, often used to aid a practitioner in the form of a computer program. The idea that these two ought to be synthesised has intuitive and empirical (Makridakis, 1988), although some note (Dawes et al., 1989) that clear problems arise in instances where the two disagree. Research does note advantages to combining the two; for instance that clinicians’ insights combined with actuarial processing can produce more accurate predictions about an outcome (e.g. Lawrence et al., 2006). A precursor of sorts to this is Goldberg’s (1970) study that involved transforming the decision processes of a clinician into a regression, finding that this regression was both more consistent and accurate than the clinician.

Whilst there is a vast literature on expert systems, much of what psychology contributes to expert systems is in the form of knowledge acquisition, and not expert systems for psychologists. There have been propositions of expert systems for areas of interest to a clinical psychologist, but these are relatively few in number. In similar domains, expert systems have been applied in medicine for physical illness and clinical trials. For example in psychiatry, expert systems have been proposed for a range of mental illnesses, such as in schizophrenia in which recent research (Sabeti et al., 2011) has outlined how an expert system which analyses EEG patterns can be used to reliably diagnose schizophrenia.

An expert system is considerably more sophisticated than an actuarial equation because it is intended to encapsulate the schemas, domain knowledge and reasoning processes that an expert utilises (Nurius & Nicoll, 1992), rather than a mathematical rule, so they can be used by a layperson. Nurius & Nicoll (1992) note the difficulty of this because such information is procedural rather than declarative; as such simply asking the expert to go through the decision processes that they make is problematic. Critically important in this instance is the selection of the expert/s from which knowledge will be acquired, for use as the basis for the expert system. Research has
investigated this, noting a number of important facets. These have tended to include high levels of experience and heterogeneity (Osuagwu & Okafor, 2010), or using a specific methodology for identification (Shanteau et al., 2002).

Some research has been carried out in some cases to develop expert systems for use in clinical prediction. Van Aarle & van der Bercken (1999) for instance outline an expert program for language difficulties, in which experts were asked to complete problems related to various difficulties.

Within clinical psychology, the area with largest amount of research is in the cessation of addictive behaviours, in particular smoking (e.g. Velicer et al., 1993; Newman et al., 2011). Other complex systems have been produced for behaviours of interest to psychologists such as Lifenet (Ferns, 1995) for suicidal behaviour. Research using psychological diagnosis has been highly limited however; systems have been produced to Diagnose personality disorders using the MMPI (Nord & Horn Nord, 1996). More recently, psychotherapists have moved towards using similar systems to add statistical processing into appraising art therapy in facets such as colour (Kim, 2010; Kim et al., 2006; 2011), although some have considered the use of these controversial (Mattson, 2010), and reported kappa values for reliability between users raise questions (Kim, 2010). Whilst the figures generally indicated very good reliability (all except a couple fell above .8), it is relatively low for an expert system.

What must be stressed from all of these is that expert systems act as a compliment rather than a replacement. Expert systems, like actuarial rules or structured judgements do not provide very high, but not perfect accuracy; a case in point is Sabeti and colleagues (2011), whose system has an accuracy of between 80 and 90%. Whilst highly accurate; Grove and colleagues (2000) found expert and actuarial accuracies around 50 and 60% respectively, it is a different type of prediction than those made by a psychologist, so this must be taken into consideration.

Other issues that are of importance are more humanistic in nature; Osuagwu & Okafor (2010) note the unease of placing human lives in the hands of a computer. This is not an objectionable fear; however much research has demonstrated the fallibility of humans, even experts, there is unease about placing decisions about health and finance in the hands of a non human (e.g. Thornett, 2001), even though computers used to aid decisions from day to day. Other research (Will, 1992) has noted that
experts using such systems take a negative view of these, and become more anxious when using them.

In terms of future research, there are many avenues with much to be discovered. Garb (2005) notes the usefulness of capturing the cognitive processes of an expert clinician, although research to achieve this appears to be lacking. Moreover whilst there are limited systems for specific psychometric tests or in some cases disorders, systems investigating disorders pertinent to psychologists have not been studied. What is obvious from current research is that the methods used in mental health involve recognition (of pattern or colour). Pattern recognition may be of use to psychologists, such as reading patterns from psychometric tests; reviews of actuarial methods (Garb, 2000) have noted that expert neuropsychologists are adept at this.

However, what would be of interest is moving towards more sophisticated systems, similar to the Multicriteria Decision Support Systems outlined by Siskos and Spyridakos (1999). Some systems have been produced similar to this (i.e. Lifenet), but these are fairly sparse. The considerable absence of diagnosis systems also presents a rich avenue for research.

Other research areas of importance would include making expert systems accessible for experts. This appears to be an issue, and like actuarial means the development of these techniques is somewhat futile if clinical opposition is strong. Moreover, as the research in art therapy has demonstrated, further work ought to be carried out to improve the inter rater reliability of these systems.
Other improvements have concerned another means of combining expert judgemental and statistical means. This can encompass a number of different techniques. Swets and colleagues (2000) for example note that judgemental and actuarial predictions can be made, with the latter acting as a second opinion, in a manner similar to the expert system. To this end some, such as Lawrence and colleagues (2006), consider the use of expert adjustments. This consists of a statistical forecast, which is amended by an expert to account for factors that the model cannot account for; such as domain knowledge, or Meehl’s (1957) broken leg example, in which a model can be highly accurate in prediction, but not account for highly improbable events in the manner an expert can. Research however has produced mixed findings. Lawrence and colleagues note that the adjustment consists of two phases, deciding whether an adjustment is needed, and the magnitude of the adjustment. Research indicated that whilst these adjustments are relatively accurate (Lawrence et al., 2006), there are issues to be aware of. For instance, Armstrong (2006) found that whilst evidence suggested the method was promising, minor adjustments introduced systematic bias. Other general issues from similar judgemental forecasting include handling of noise where considerable noise is present (Harvey, 1995; Lawrence et al., 2006). Other biases such as anchoring, overreaction/optimism (Eroglu & Croxton, 2010; see Lichtenstein & Fischoff, 1977 for examples) and confirmation bias (Soll & Mannes, 2011; see Nickerson, 1998 for a review) have also been identified in adjustments. The presence of confirmation bias is particularly important as it mirrors strongly the deficiencies in prediction that have led research to promote actuarial processing of clinical prediction (Chapman & Chapman, 1968). Others (Swets et al., 2000; Grove & Meehl, 1996) have noted the situations under which adjustments would be appropriate. These focus primarily upon extenuating factors which are highly pertinent to the criterion being measured, but not a factor usually encountered. The other listed is populations which may not be accounted for by current research. Evidence suggests that the efficacy of this methodology is somewhat mixed; although research evidence has demonstrated that expert adjustments appear to be quite efficacious, research is relatively emergent.
in comparison to other forecasting techniques (Armstrong, 2006), and there remains the issue of individual biases present in forecasters.

STRUCTURED JUDGEMENT

Whilst considerable research has argued over the extent to which clinical predictions are valid (see Meehl, 1986; Grove & Meehl, 1996), there is a consensus in researching prediction on the caution that should be applied to unstructured clinical predictions (e.g. Hanson & Thornton, 2000; Armstrong, 2006). To this end, research has also investigated structured judgements. In clinical psychology these are often interview guidelines, which attempt to increase reliability between clinicians by ensuring that all of the criteria that define a mental illness (using the DSM-IV or ICD 10 for example) are included, whilst still the clinician room to probe areas they may see as pertinent. Beyond the support for structured expert judgement in the forecasting literature, there is evidence for the importance of utilising this in clinical psychology; research (Garb, 2005) has noted that in the process of gathering data in making a prediction, practitioners vary considerably in their evaluation of weightings relevant to evaluating the criteria for a mental disorder. Other research (Bloom, 1992) has noted this as a particular advantage of computer over clinician. Moreover structured interviews appear to lead to more diagnoses with poor concordance with clinical methods (Shear et al., 2000), although it remains unclear whether this is a result of greater sensitivity or a higher false positive rate. It appears a large number fail to ask questions that delineate mental illnesses. As such, introducing elements of structure ensure that the necessary amount of pertinent data ought to improve clinical accuracy by encouraging clinicians to gather a sufficient amount of data to enable them to make an accurate prediction. Garb (2005) does however note that caveats that must be taken with such an approach. Most notably, clinical judgement must be exercised to note the situations under which alternative diagnostic methods are optimal. There are however issues to be aware of. Wood and colleagues (2002), in their review of clinical assessment, note that the changing ethnic memberships of the population (although most pertinent to the US, not irrelevant in the UK) mean that clinical instruments must be sensitive to differences in this respect. This highlights an
important issue within structured decision aids, which is the role of the clinician’s judgement.

CONCLUSION

There are a number of means to improve clinical prediction. The means discussed in this paper, following on from the improvements recommended by forecasting research, generally work by adding structure to the forecasting process of the practitioner. These work at all of the points of the reasoning and forecasting process; data collection, statistical prediction and criteria from said statistical reasoning what state of affairs best represents the patient. Researchers however have stressed these aren’t perfect, and are intended to aid rather than replace the decision maker (except in the case of statistical reasoning). A distinction must also be made between the types of prediction psychologists make: there are the broad profiles that make up a disorder or pathology, but also specific behaviours such as self-harm and suicide. As discussed, some of the proposed means of improvement are much more useful in improving clinician’s predictions of broad profiles than individual behavioural outcomes. Other research has pointed out the opposite. One possible explanation for this is that actuarial means are simply better where theoretical knowledge about a disorder or maladaptive behaviour is well developed. It would stand to reason therefore if our understanding of mental health increases, in turn so will the means of improving clinical judgement. A number of means have been proposed, be it actuarial rules, structured judgements, judgemental adjustments and expert systems. These all can work in tandem with one another, with actuarial rules and judgemental adjustments being a powerful example of this.
REFERENCES


