Article:
Michael, B., Kellett, S. and Delgadillo, J. (2023) Is clinical decision-making in stepped-care psychological services influenced by heuristics and biases? Behavioural and Cognitive Psychotherapy, 51 (4). pp. 362-373. ISSN: 1352-4658

https://doi.org/10.1017/S1352465823000115

Is clinical decision-making in stepped-care psychological services influenced by heuristics and biases?

Benjamin Michael[1*], Stephen Kellett[2,3], & Jaime Delgadillo,[3,2]

1. Leeds and York Partnership NHS Foundation Trust, United Kingdom
2. Rotherham Doncaster and South Humber NHS Foundation Trust, United Kingdom
3. Clinical and Applied Psychology Unit, Department of Psychology, University of Sheffield, United Kingdom

Conflicts of interest: None.

Corresponding author: Dr. Benjamin Michael, benjamin.michael@nhs.net

Abstract

Background. The manner in which heuristics and biases influence clinical decision-making has not been fully investigated and the methods previously used have been rudimentary.

Aims. Two studies were conducted to design and test a trial-based methodology to assess the influence of heuristics and biases. Specifically, with a focus on how practitioners make decisions about suitability for therapy, treatment fidelity and treatment continuation in psychological services.

Method. Study one (N=12) used a qualitative design to develop two clinical vignette-based tasks that had the aim of triggering heuristics and biases during clinical decision making. Study two (N=133) then used a randomised crossover experimental design and involved psychological wellbeing practitioners (PWPs) working in the Improving Access to Psychological Therapies (IAPT) programme in England. Vignettes evoked heuristics (anchoring and halo effects) and biased responses away from normative decisions. Participants completed validated measures of decision-making style. The two decision-making tasks from the vignettes yielded a clinical decision score (CDS; higher scores being more consistent with normative/unbiased decisions).

Results. Experimental manipulations used to evoke heuristics did not significantly bias CDS. Decision-making style was not consistently associated with CDS. Clinical decisions were generally normative, although with some variability.

Conclusions. Clinical decision-making can be “noisy” (i.e., variable across practitioners and occasions), but there was little evidence that this variability was systematically influenced by anchoring and halo effects in a stepped-care context.

Introduction

The Improving Access to Psychological Therapies (IAPT) programme in England is a national programme offering rapid access to evidence-based psychological therapies recommended by clinical guidelines for the treatment of anxiety and depressive disorders (National Institute for Health and Care Excellence [NICE], 2011). IAPT services follow a stepped-care model, where many patients initially access a low-cost and brief intervention, followed by more intensive psychotherapies for those who do not fully benefit from the initial steps of treatment (Bower & Gilbody, 2005). Over one million patients per year are referred to IAPT services and the majority of patients only access low intensity psychological interventions (Clark, 2019). Therefore, the organizational efficiency and clinical effectiveness of the overall system is heavily dependent on the assessment and treatment skills of the practitioners working in the early steps of IAPT services.

In IAPT, the screenings for the service are conducted by Psychological Wellbeing Practitioners (PWPs). PWPs are trained via a national curriculum (UCL, 2015) to assess the suitability of patients with mild-to-moderate anxiety and depression for brief psychoeducational and low intensity interventions. Saffron et al., (2021) defined one-to-one interventions delivered by PWPs as based on a self-help approach supported by didactic materials, where patients typically receive up to six hours of contact time, with sessions being 30-minutes each. Patients are therefore 'stepped-up' to more intensive therapies according to need, risk, and lack of responsivity to any initial low intensity approach. The role of the PWP was established with the implementation of the IAPT programme and workforce (Kellett et al., 2020). Despite the fact that the PWP workforce is widely available across England, contributing to the assessment and treatment of thousands of patients per year, relatively little research is available on clinical decision-making by PWPs.

Clinical judgement is known to be influenced by biases, which results in relatively poor agreement between psychological professionals and also variability in the quality and accuracy of clinical decisions (Garb, 2005; Grove, 2005; Grove & Meehl, 1996). When making diagnostic or prognostic assessments clinical judgment has been found to be less accurate compared to structured algorithms or statistical models (Ægisdóttir et al., 2006). Clinical decision-making is influenced by a myriad of variables such as the patient's characteristics, attitudes, preferences, interpersonal relationships, and the confidence/competence of the practitioner (Stavrou et al., 2009; Anthony et al., 2010; Sigel & Leiper, 2004; Pilgrim et al., 1997; Visintini et al., 2007). In an IAPT context, Delgadillo el al. (2015) investigated 'stepping decisions' and found four factors were associated with offering longer treatments for unresponsive patients; (a) when there were obstacles in stepping up or referring on; (b) when the client was liked by the therapist; (c) if there was a positive therapeutic alliance; (d) when a positive outcome was envisaged through extending treatment. In this study, the retention of a patient in a treatment that was not resulting in reliable improvements was considered to be an obstacle to the adequate functioning of a stepped care treatment system, where other (more intensive) treatments could be offered instead. Delgadillo et al. (2015) therefore concluded that incongruence and inaccuracy in decision-making were due to a complex interplay of beliefs, attitudes, subjective norms, and perceptions of self-efficacy. Despite the complex interplay of variables that impinge on clinical judgment, it is likely that psychological professionals integrate this information in an intuitive and "fast" way often using mental shortcuts. In his seminal work on decision-making, Kahneman (2011) proposed that rather than combining multiple sources of information using complex mental processes to make decisions, people often rely on "system 1" (i.e., "fast" thinking), which uses heuristics and biases (Kahneman & Tversky, 1972).

Heuristics (i.e., unconscious mental shortcuts) and cognitive biases (i.e., systematic tendency to jump to certain conclusions) influence how people make intuitive decisions in daily life (Tversky & Kahneman, 1974). Tversky and Kahneman (1974) introduced two cognitive processes that appear important when considering clinical decision-making: (a) *'anchoring and adjustment'* whereby excessive significance is placed on the first piece of information encountered when making a decision; and (b) the *'halo effect'* whereby an impression formed from a single characteristic is allowed to influence multiple judgments of unrelated factors. Cognitive processes such as those described above have been proposed to influence decisions in routine clinical care

(Garb, 2005), ultimately having an impact on patients' treatment pathway, experiences, and outcomes.

The most common method of investigating the influence of heuristics and biases is using case vignettes requiring participants to make judgments and decisions when knowing the 'correct' course of action (e.g., Spengler & Strohmer, 1994; Garb, 1996; Berman et al., 2016). However, the reliability and ecological validity of using generic case vignettes has been questioned (e.g., Hyler, Williams, & Spitzer, 1982). Therefore, it is necessary to develop vignettes that have face-validity with the professional groups being tested and the decisions that encountered in routine practice. Given the scarcity of research related to clinical decision-making in stepped-care psychological services, the present study developed and applied a case vignette methodology to study the potential influence of heuristics and biases in a PWP sample. This is because of the acknowledged role of PWPs in screening referrals to the IAPT programme. We were interested in investigating whether cognitive processes such as the *anchoring* and *halo* effects may influence clinical decisions concerning treatment suitability (i.e., if therapy is an appropriate option), treatment fidelity (i.e., delivering a protocol-driven intervention) and treatment continuation (i.e., decision to lengthen the duration of a treatment). These aspects of decision-making are critical for the efficient and effective use of stepped care interventions (NICE, 2011). Two linked studies were conducted. Study one used a qualitative design to develop an ecologically valid, clinical vignette-based method to assess the influence of heuristics and biases on clinical decision making. The second study used a randomised crossover experimental design using the clinical vignettes to examine the influence of heuristics and biases in a sample of PWPs. We hypothesized that evoking heuristics and biases would prime counter-normative decisions (i.e., as opposed to normative and clinical-guideline adherent decisions). We also hypothesized that respondents' general decision-making style (rational vs. intuitive) would be significantly associated with clinical decision-making.

Method

Ethical approval was granted by the relevant university Ethics Committee (ref: 017478). Informed consent to participate and for the results to be published was obtained. Participants' right to privacy was also respected. Authors abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the BABCP and BPS.

Study A

Development of a clinical case vignette methodology

Previous studies using case vignette methods prompted respondents to choose either "normative" (i.e., logical/expected) or "counter-normative" (i.e., intuitive/biased) choices/decisions (e.g., Kahneman & Tversky, 1972; Tversky & Kahneman, 1974). Following this paradigm, we developed case vignettes that would have ecological validity to typical decisions encountered by PWPs in stepped care services. To design the vignettes, an inductive process was undertaken informed by ethnographic decision tree modelling (EDTM; Gladwin, 1989), which included 8-steps in the development of a composite group model. The structure of the decision-making task and the scoring system are presented in Figure 1.

Figure 1 about here

Participants and processes for focus group and piloting

The development process included three phases: [1] A preliminary draft of clinical case vignettes was developed by the research team, informed by the literature on heuristics and biases, and with a specific focus on anchoring and halo effects. [2] Two experienced PWP educators were recruited to contribute to a consultation focus group, to check the face validity of proposed content of the clinical case vignettes. A semi-structured interview document was developed to guide the focus group and to acquire qualitative data relating to the structure and content of the measure. The focus group lasted one hour, and discussions were audio-recorded and fully transcribed. A 'living document' (Shanahan, 2015) was then developed and was reviewed by research supervisors until consensus was reached regarding content. [3] Then, N=10 PWP educators were recruited for the piloting stage. Piloting entailed participants engaging in the simulation of the full study (see procedure listed in Study B) and providing feedback. Feedback from the pilot study was then incorporated into the final case vignette design.

Analysis strategy

Data from the focus group were transcribed verbatim and analysed using the six-phases of thematic analysis (Braun & Clarke, 2006). A secondary coder (trainee clinical psychologist) independently reviewed preliminary codes, themes, and sub-themes. The percentage agreement score was 97.8% and Krippendorff's alpha was 0.79. The resulting case vignettes and experimental manipulations are described below.

Study B

Design and Participants

A randomized crossover experimental study designed to evoke specific cognitive processes (i.e., anchoring and halo effects) and to measure their potential influence on clinical decision-making. Participants were recruited using a convenience sampling method from a national workforce of trainee and qualified PWPs working clinically within the IAPT programme (Clark et al., 2009). Recruitment took place via email by approaching PWPs via the Psychological Professions Network, Health Education England, British Psychological Society PWP training committee, and PWP course directors network list (nationally). Consenting participants received an information sheet and those who provided consent (using an electronic survey) were consecutively included in the experiment during a four-month study period.

Procedure: experimental manipulation

The experiment was conducted online using the Qualtrics platform. The study followed ethical guidelines for internet mediated studies (British Psychological Society, 2017). Participants were required to read and work through two case vignettes (named “Jack” and “Chloe”), each of which prompted them to record the decisions they would make about each patient’s treatment. The experiment was designed in such a way that one of these vignettes contained an experimental manipulation designed to evoke anchoring and halo effects, while the other vignette served as a control condition. Since participants completed two tasks, one of which was a control condition, the order in which they completed each scenario was decided by randomization. Furthermore, to control for the influence of spurious details of each case vignette (i.e., word count, gender of the patient, etc.), the inclusion of the experimental manipulation was also decided by randomization. For instance, if a participant was randomized to the sequence Jack-Chloe, where the experimental manipulation was randomized to the Chloe vignette, they would complete the control version of the Jack vignette.

Clinical case vignette methodology

Each case vignette contained a brief description of a patient and presented the participant with three clinical decision-points. These decision-points reflected common questions arising in routine care relating to [1] *suitability* (i.e., is this patient suitable for therapy?), [2] *treatment fidelity* (i.e., should I continue to offer a standard and protocol-driven treatment?), and [3] *treatment continuation* (i.e., should I continue to offer treatment A, or should I refer the patient to treatment B?). At each decision-point, participants were asked to choose the statement (from a list of options) that best resembled what they would decide in routine care. Options were conceptualised a priori as either “normative” (i.e., following clinical guidelines) or “counter-normative” (i.e., deviating from clinical guidelines).

The experimental version of the vignettes was designed to evoke/prime heuristics and biases, in a way that might increase the likelihood of counter-normative responding. Informed by previous research in the context of stepped care psychological services (Delgadillo et al., 2015), we hypothesised that emotionally evocative patient-features (i.e., complicated, highly distressed, potentially difficult to work with) would increase the likelihood of counter-normative (i.e., improvisational and intuitive) decisions which would deviate from those recommended by clinical guidelines. Such features were presented early in the case vignette, expecting that respondents may become influenced by their "first impression" (i.e., the anchor) and then may make later decisions with reference to it (i.e., the halo effect). As each case vignette required three decisions, each "normative" decision was coded "1" and each counter-normative decision was coded "0". Hence, each case vignette yielded a 0-3 clinical decision score (CDS), where higher scores denoted a greater propensity to normative decision-making. The CDS was the primary outcome measure.

Measures

Participants completed validated measures of decision-making style, reflective capacity, and personality. The sequence in which questionnaires were presented to each participant was decided by randomization. *Cognitive Reflection Test (CRT; Frederick 2005).* The CRT is a three-item measure that measures the tendency to override an initial "gut" response and engage in further reflection to find a correct answer. It has been shown to account for a substantial unique variance (11.2%, $p < .001$) in decision-making choices after other measures of individual differences are statistically controlled. *Rational and Intuitive Decision Styles Scale (DSS; Hamilton, Shih & Mohammed, 2016).* This is a 10-item decision style scale capturing a broad range of the rational/intuitive thinking styles construct domains. Test–retest reliability has been reported to be high for both rational ($r = .79$, $p < .01$) and intuitive ($r = .79$, $p < .01$) dimensions. The DSS has demonstrated high internal consistency and a robust two-factor structure.

Statistical analysis

Sample size calculation. Toplak (2011) demonstrated that the Cognitive Reflection Test (CRT) shows an average correlation of r=.49 with performance on heuristic and biases tasks (i.e., decision-making case vignettes). According to Cohen's (1977) method a sample size of N=38 participants per group would be necessary to detect a large effect size ($r = 0.50$), with an alpha or significance level of 0.05, and 80% power in an experimental design. This yielded a minimum sample size requirement of N=76.

Primary analysis. First, Kruskal-Wallis Tests were used to examine differences in mean CDS, comparing experimental vs. control tasks. These tests were computed twice, once for each of the case vignettes (i.e., "Jack" and "Chloe"), enabling us to examine the overall influence of the experimental manipulation across all clinical decisions featured in these scenarios.

Secondary analyses. Next, linear regression analyses were used to examine if the experimental manipulation (independent variable) influenced CDS (dependent variable) after controlling for individual differences in general decision-making style (CRT, DSS rational/intuitive subscales). Separate regressions were computed for each case vignette. Finally, logistic regressions controlling for general decision-making style (as above) were used to assess if the experimental manipulation was associated with a higher probability of counter-normative decisions for the specific tasks relating to the second (*treatment fidelity*) and third (*treatment continuation*) decision-points in each vignette. Since the first decision-point related to a patient's suitability for treatment, and all participants decided the patients were suitable, there was no variability in the first decision-point. Therefore, the logistic regressions were only applied to the second and third decision-points.

Results

Study one

Experimental and control versions of each of the two clinical case vignettes were co-produced through development, consultation, and piloting phases. These case vignettes and further information relating to their development are available in supplemental materials. Sample characteristics were not collected in this small sample in order to safeguard anonymity. The vignettes were deemed suitable to then progress onto Study Two.

Study two

Sample characteristics. N=133 participants who completed the experimental task were included (excluding N=57 who consented but did not complete the task). The participants worked across 16 counties in England. The majority described their gender as female (86.3%). The mean age was 32.86

years (SD = 9.13). A large proportion were either qualified PWPs (54.9%) or senior PWPs (19.6%), with an average 5-years of clinical experience in the PWP role (M = 5.02; SD = 3.37).

Primary analysis. Kruskal-Wallis Tests indicated no significant differences in mean CDS comparing experimental vs. control conditions in either of the two clinical case vignettes.

Secondary analyses. Linear regression results are reported in Table 1. The regression coefficient for the experimental manipulation was not statistically significant in either model. The only significant predictor in the model was the DSS rational subscale score, which was positive correlated with CDS; $\beta = .19$, $p < .05$.

[Table 1]

Logistic regression results are presented in Tables 2 and 3. One of the two logistic regressions (case vignette "Chloe") indicated that the experimental manipulation was significantly associated with a higher probability of counter-normative decisions regarding treatment fidelity. However, this did not replicate in the other case vignette ("Jack"). One of the two logistic regressions (case vignette "Jack") indicated that the experimental manipulation was significantly associated with a higher probability of normative decisions regarding treatment continuation. However, this did not replicate in the other case vignette ("Chloe").

[Tables 2 and 3]

Discussion

This study investigated clinical decision-making by PWPs working within IAPT stepped-care services in England. Decision-making by PWPs has not been previously investigated and therefore the study was novel. The study examined whether specific cognitive processes (anchoring and halo effects) and more general decision-making styles (i.e., rational vs. intuitive) influenced care decisions commonly encountered in stepped-care clinical practice. To this end, we first produced clinical case vignettes in Study One that had face validity with PWPs, which resembled realistic clinical cases, and so required PWPs to consider typical dilemmas relating to evidence-based practice. In Study Two, we conducted a cross-over randomised control trial and found that clinical care decisions were not significantly influenced by anchoring and halo effects or the general decision-making style (CRT and DSS scales) of PWPs. In short, the decisions made regarding suitability, treatment fidelity and treatment continuation made by practicing PWPs were not systematically biased by anchoring and halo effects.

It is noteworthy that, in one of the vignettes, the apparent effect of the experimental manipulation was contrary to our predictions, since it apparently increased normative rather than counter-normative decisions. This may be because all the sample in the experiment were practicing PWPs and therefore would have weekly case management supervision (UCL, 2015). In this type of supervision, PWPs are obviously encouraged and so practice making rational decisions concerning patient care. This practice effect may therefore have dampened the influence of the anchoring and halos implemented in the experiment. The normative versus counter-normative ratios suggests that this spurious result may have been also influenced by "noise": variability due to haphazard aspects of the case vignette (e.g., gender, word count, etc.) or occasion (e.g., respondent fatigue, distraction, etc.), rather than the experimental manipulation itself. It is plausible that the case vignettes and priming tasks did not optimally capture the natural variability that may better characterise decision-making in routine care. It is likely that there is considerable variability in decisions not only between PWPs but also within PWPs (i.e., over time and across multiple cases). This natural variability or "noise" in professional judgments has been previously documented across various occupations (see Kahneman et al., 2021), and can be modelled by sampling multiple respondents' judgements across multiple cases (rather than only one or two cases). Data from the present study reconfirmed that clinical decisions can vary from one respondent to another (i.e., some made normative, and others made counter-normative choices). This is consistent with data from previous studies that show wide variability in actual treatment selection decisions made by several psychological professionals across multiple cases in routine practice (e.g., Delgadillo et al., 2017).

Strengths, limitations and future research

This study followed a rigorous, theoretically-informed and multi-phase method to design ecologically valid and realistic clinical case vignettes. This process involved PWPs working in the relevant clinical setting and recruited from nation-wide mailing lists. The experimental design had built-in controls for ordering effects and was sufficiently powered; the actual sample (N=133) was nearly twice as large as the minimum requirement (N=76).

But the study also had several limitations that inform the design of future studies. The convenience sampling approach used to recruit participants could be affected by self-selection bias. For example, very busy PWPs may have declined to participate, potentially limiting the ecological representativeness of the sample. The study only collected basic participant characteristics, and these were not linked to data from the experimental task to safeguard anonymity and to promote participation. Therefore, it was not possible to assess if these or other unmeasured features may have influenced decision-making during experimental tasks. The CDS metric had a narrow range (0-3), which may have artificially constrained variability.

An analogue approach was employed, rather than studying decision-making in a naturalistic setting. Whilst strengths to the analogue approach include tighter control of variables, it is acknowledged that participants might have been inclined to respond in a socially desirable manner (Hare-Mustin, 1983). Emotional evocation in case vignettes is likely to be less intense than simulated (e.g., role play) or actual clinical encounters. Participants may have felt less connected and empathic towards the analogue patients than ‘real’ patients and this may have increased the likelihood of providing more normative responses. Results, therefore, may not be a true reflection of how PWPs respond in a real-life clinical setting. Future research should therefore seek to address limitations regarding the design of the dynamic measure such as the lack of variability in the scoring system and potential limitations regarding its ecological validity. Employing fictional video-recorded clinical scenarios rather than case vignettes whilst using a forced choice format to control for variability could be a methodological improvement in future studies. Comparing decision-making styles of participants according to their level of experience could also be appropriate.

Prior to this study, clinical decision making of psychological professionals has not been widely investigated. Therefore, previous research enabling comparison with the findings from this study were not available. Furthermore, literature related to clinical decision making more generally and the use of the case vignette method in research has not been updated for over 10 years. This is a drawback for the present study, especially compared to the relative newness of the PWP role focused on here (UCL, 2015). Another limitation relates to the fact that there is no explicit reference to risk with case vignette 'Jack', yet risk is more explicit with case vignette ‘Chloe’. This difference was not considered within the analysis. Furthermore, there is frequent reference to ‘Jack’s’ feelings of hopelessness without any specific risk assessment information. This may have influenced clinical decision making in this case vignette. This is a limitation given hopelessness is a key area of risk assessment in clinical practice (Ribeiro et al., 2018) and therefore likely to be picked up on by participants. Whilst the study design was adequate to study systematic bias, the limited number of vignettes presented to each participant did not allow us to adequately model the true extent to which clinical decisions may be “noisy”.

Conclusion

Making clinical decisions is a key part of any clinical role and making rational decisions is in the best interest of the patient. PWPs due to their place in stepped-care IAPT services make many decisions about care to be delivered and therefore ensuring lack of obvious bias in these decisions is important. Clinical decisions can vary between practitioners who encounter similar clinical scenarios. Often, practitioners make decisions that accord with clinical guidelines, but sometimes decisions can be counter normative. We found no convincing evidence that variability in decision-making was systematically biased by anchoring and halo effects, but in a sample exposed to a regular check on decisions made (i.e., case management supervision). Variability in clinical decisions observed in routine care and in analogue tasks may be better explained by "noise": individual differences and natural fluctuations in judgment quality.

Data availability statement: In line with the requirements of the ethics review board for this study, requests for access to data are to be made in writing to the corresponding author.

References

Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A., Cook, R.S., Nichols, C.N., Lampropoulos, G.K., Walker, B.S., Cohen, G. & Rush, J.D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*, 341-382. https://doi.org/10.1177/0011000005285875

Anthony, J. S., Baik, S. Y., Bowers, B., Tidjani, B., Jacobson, C. J. and Susman, J. (2010). Conditions that influence a primary care clinician's decision to refer patients for depression care. *Rehabilitation Nursing, 35,* 113–122. https://doi.org/10.1002/j.2048-7940.2010.tb00286.x

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: access, effectiveness, and efficiency: narrative literature review. *The British Journal of Psychiatry, 186*, 11-17. https://doi.org/10.1192/bjp.186.1.11

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*, 77-101. https://doi.org/10.1191/1478088706qp063oa

Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R. & Wright, B. (2009). Improving access to psychological therapy: initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, *47*, 910–920. https://doi.org/10.1016/j.brat.2009.07.010

Clark, D.M., (2019). *NHS England: IAPT at 10: Achievements and challenges.* Retrieved from: https://www.england.nhs.uk/blog/iapt-at-10-achievements-and-challenges/

Cohen, J. (1977). *Statistical power analysis for the behavioural sciences* (revised ed.).

Delgadillo, J., Gellatly, J., & Stephenson-Bellwood, S. (2015). Decision making in stepped care: how do therapists decide whether to prolong treatment or not? *Behavioural and cognitive psychotherapy*, *43*, 328-341. https://doi.org/10.1017/S135246581300091X

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of consulting and clinical psychology, 85*(9), 835–853. https://doi.org/10.1037/ccp0000231

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*, 25-42.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67-89. https://doi.org/10.1146/annurev.clinpsy.1.102803.143810

Gladwin, C. H. (1989). *Ethnographic decision tree modelling*. (Vol. 19). Sage.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323. https://doi.org/10.1037/1076-8971.2.2.293

Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of clinical psychology*, *61*, 1233-1243. https://doi.org/10.1002/jclp.20179

Hamilton, K., Shih, S. I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of personality assessment*, *98*, 523-535. https://doi.org/10.1080/00223891.2015.1132426

Hyler, S. E., Williams, J. B., & Spitzer, R. L. (1982). Reliability in the DSM-III field trials: interview v case summary. *Archives of General Psychiatry, 39*, 1275-1278. https://doi.org/10.1001/archpsyc.1982.04290110035006

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, *3*, 430-454. https://doi.org/10.1016/0010-

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: a flaw in human judgment*. London: William Collins.

National Institute for Health and Care Excellence (2011). *Common mental health problems: identification and pathways to care [CG123].* Retrieved from: https://www.nice.org.uk/guidance/CG123/chapter/1-Guidance#stepped-care

Pilgrim, D., Rogers, A., Clarke, S., & Clark, W. (1997). Entering psychological treatment: decision-making factors for GPs and service users. *Journal of Interprofessional care, 11*, 313-323. https://doi.org/10.3109/13561829709034128

Ribeiro J. D., Huang, X., Fox, K. R., & Franklin, J. C. (2018). Depression and hopelessness as risk factors for suicide ideation, attempts and death: meta-analysis of longitudinal studies. *British Journal of Psychiatry, 212(5),* 279-286. doi: 10.1192/bjp.2018.27.

Qualtrics (2002). Retrieved from: https://www.qualtrics.com/uk

Shafran R., Myles-Hooton P., Bennett S, & Öst L.G. (2021). The concept and definition of low intensity cognitive behaviour therapy. *Behavour Research and Therapy, 138*, 103803. doi: 10.1016/j.brat.2021.103803.

Shanahan, D. R. (2015). A living document: reincarnating the research article. *Trials, 16*, 151. https://doi.org/10.1186/s13063-015-0666-5

Sigel, P., & Leiper, R. (2004). GP views of their management and referral of psychological problems: a qualitative study. Psychology and Psychotherapy: Theory, Research and Practice, 77(3), 279-295. https://doi.org/10.1348/1476083041839394

Stavrou, S., Cape, J., & Barker, C. (2009). Decisions about referrals for psychological therapies: a matched-patient qualitative study. *Br J Gen Pract, 59*, e289-e298. https://doi.org/10.3399/bjgp09X454089

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition, 39*, 1275.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185,* 1124–1130. https://doi.org/10.1126/science.185.4157.1124.

UCL (2015). *National Curriculum for the Education of Psychological Well-being Practitioners, 3rd edn*. Available at: https://www.ucl.ac.uk/pals/research/cehp/research-groups/core/pwp-review/docs/PWPREVIE_-curriculum

Visintini, R., Ubbiali, A., Donati, D., Chiorri, C., & Maffei, C. (2007). Referral to group psychotherapy: A retrospective study on patients' personality features associated with clinicians' judgments. *International Journal of Group Psychotherapy, 57*, 515-524. https://doi.org/10.1521/ijgp.2007.57.4.515

Table 1. Summary of Multiple Regression Analyses for Variables Predicting Clinical Decision Scores (CDS) in the Case vignette (CV) 1 and 2 experiments (N = 131)

	CV1			CV2		
Variable	*B*	*SE B*	β	*B*	*SE B*	β
Group (Exp/Con)	-0.08	0.04	-.17	0.08	0.05	.15
CRT Measure	-0.02	0.02	-.11	-0.01	0.02	-.02
DSS Rational Subscale	0.02	0.01	.19*	-0.01	0.01	-.08
DSS Intuitive Subscale	0.01	0.01	.10	-0.01	0.01	-.07
Adjusted R^2		.04			-.00	
F		2.35			0.98	

*$p < .05$; *B* = unstandardized regression coefficient; *SE B* = standard error of the coefficient; β = standardized coefficient

Table 2. Logistic Regression Predicting the Treatment Fidelity and Step Up/Treatment Continuation Questions in the CV1 Experiment[1]

Question		*B*	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio	
								Lower	Upper
Treatment Fidelity	Group (Exp/Cont)	-19.54	.4585.816	.00	1	.997	.000	.000	.
	CRT Measure	.693	.380	3.32	1	.068	2.00	.95	4.22
	DSS Rational Subscale	-.25	.21	1.40	1	236	.78	.52	1.18
	DSS Intuitive Subscale	.22	.16	2.09	1	.149	1.25	.92	1.70
	Constant	22.94	4585.82	.00	1	.996	9210021747.15		
Step Up/ Continue	Group (Exp/Cont)	1.43	.39	13.20	1	<.001	4.19	1.93	9.06
	CRT Measure	.16	.16	1.04	1	.309	1.17	.86	1.59
	DSS Rational Subscale	-.161	.08	3.80	1	.051	.85	.72	1.00
	DSS Intuitive Subscale	-.096	.07	1.93	1	.165	.91	.79	1.04
	Constant	4.12	2.23	3.41	1	.07	61.45		

[1] In both stages of the clinical decision-making task (treatment fidelity and hold/step up) for both vignettes (1 and 2) the reference category (0) was the control group, and the signal category (1) was the experimental group. The dependent variable was coded “1” for a normative answer and “0” for a counter-normative answer.

Table 3. Logistic Regression Predicting the Treatment Fidelity and Step Up/Treatment Continuation Questions in the CV2 Experiment[2]

Question		*B*	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio Lower	Upper
Treatment Fidelity	Group (Exp/Cont)	-1.22	.37	10.60	1	.001	.30	.14	.62
	CRT Measure	.12	.15	.61	1	.436	1.12	.84	1.51
	DSS Rational Subscale	.02	.08	.04	1	.839	1.02	.87	1.18
	DSS Intuitive Subscale	-.002	.07	.00	1	.978	1.00	.88	1.14
	Constant	.476	2.15	.05	1	.825	1.61		
Step Up/ Continue	Group (Exp/Cont)	.474	.36	1.70	1	.194	1.61	.79	3.28
	CRT Measure	-.042	.15	.09	1	.770	.96	.72	1.28
	DSS Rational Subscale	.07	.08	.77	1	.381	1.07	.92	1.24
	DSS Intuitive Subscale	.07	.07	1.11	1	.293	1.07	.94	1.22
	Constant	-2.05	2.11	.94	1	.332	.13		

[2] In both stages of the clinical decision-making task (treatment fidelity and hold/step up) for both vignettes (1 and 2) the reference category (0) was the control group, and the signal category (1) was the experimental group. The dependent variable was coded “1” for a normative answer and “0” for a counter-normative answer.