Anxious People, Please Apply! No Evidence for Decreased Perceptions of Employability in Individuals with Mental and Physical Illness Compared to Healthy Controls

Maximin Lange & Ricardo Twumasi
King's College London

This paper builds on and updates Koser et al. (1999) and Feria et al. (2014) by investigating the degree to which individuals applying for executive, administrative and manual job positions may experience bias in selection, owing to self-reported mental and/or physical disabilities in a United Kingdom (UK) sample. Comparing the impact of different disclosed disabilities on anticipated candidate selection ratings found no evidence that people with a former mental or physical illness were rated lower than those without such conditions. Mediation analyses revealed anticipated deficits in attributes stereotypically low in mental health patients, i.e., dependability or resilience, contributed to participants' negativity in predicted competence of presented candidates. Our findings suggest that disclosing a psychiatric history of anxiety in an employment context may not impact career opportunities. However, heightened associations with mental illness stereotypes were shown to impact the professional fitness evaluations of individuals disclosing previous mental ill health.

Keywords: Anxiety disorder, employment outcomes, workers with mental illness, hiring decisions, discrimination

Anxiety disorders (AD) are the most prevalent global mental health conditions across all ages (Baxter et al., 2014; Essau et al., 2018; Kessler et al., 2005; Kessler et al., 2007; Polanczyk et al., 2015; Stein et al., 2017) with a past year prevalence rate of 6.7% (Steel et al., 2014) and a 10.4% lifetime prevalence in western cultures (Baxter et al., 2013). Based on data by the World Health Organisation's World Mental Health Survey (Kessler & Üstün, 2004; Moussavi et al., 2007), across low, middle and high-income countries, measured by means of the Sheehan Disability Score out of all mental health disorders, AD were found the most disabling for sufferers and scored even higher on the measures than some physical conditions, including cancer (Ormel et al., 2008).

Author info: Correspondence should be sent to: Dr. Ricardo Twumasi, Psychosis Studies, IoPPN, King's College London, 16 De Crespigny Park, London SE5 8AF, United Kingdom email: ricardo.twumasi@kcl.ac.uk;
North American Journal of Psychology, 2022, Vol. 24, No. 2, 319-336.

Correspondence Experiments

Causal evidence of direct discrimination in the labour market is difficult to uncover, not least since such behaviour is prohibited by law and managers therefore are hesitant to state true reasons for hiring decisions. Other approaches, e.g., employment rate comparisons or subjective perceptions do suggest labour market discrimination, however, they should be interpreted with caution due to the high number of confounders (Hipes et al., 2016). Although research on labour market discrimination against people with current or former mental health issues has been conducted, this is not in exhaustive quantity, nor is homogeneity of methods ensured. Field experiments, namely correspondence experiments using matched CV designs are widely viewed as the gold standard in providing evidence of discrimination in hiring decisions (Baert, 2018). Yet, these studies have only been conducted sparingly with regards to mental illness. In his comprehensive overview of correspondence experiments concerning hiring discrimination in the mid 2000-2010s, out of 90 studies, Baert (2018) only identified three relating to mental health. One study was concerned with depression (Baert et al., 2016), one with autism (Baert 2016) and the other with an unspecified mental health problem (Hipes et al., 2016). Earlier work, not covered in the review, was done by Pearson et al. (2003) who also focused on depression. Later correspondence experiments include Feria et al. (2014), investigating depression, Sterkens et al. (2021), who looked at former burnout patients, Ameri et al. (2017) focusing on Asperger's, and Bjørnshagen (2021) investigating an unspecified mental health issue.

Matched CV Survey Designs

Correspondence experiments are hard to gain ethical approval for, given their deceptive nature and the fact that real applicants might be disadvantaged by the research procedure, since the fictional participant might get invited instead of actual candidates. Closest to mirroring real world settings come matched CV survey designs that evaluate the perceived employability of applicants in the eyes of recruiters, hiring managers and Human Resources (HR) professionals. For depression, these have been conducted by Hazer and Bedell (2000), Glozier (1998), and Kapoor (2017). Gouvier et al. (2003), looked at an unspecified mental illness, while Bricout and Bentley (2000) focused on schizophrenia. Zissi et al. (2007) included both schizophrenia and depression. Koser et al. (1999) focused on taking medication for anxiety and depression. Dalgin and Bellini (2008) as well as Bell and Klein (2001), found null results comparing unspecified psychiatric conditions

with physical ones and looking amongst other diseases at depression, respectively.

Present Research

Only Koser et al. (1999) included anxiety in addition to discrimination of depressed participants. No study has exclusively focused on anxiety disorder of potential employees. Additionally, Glozier (1999) conducted the only study using a UK sample. Some of the studies (Gouvier et al., 2003; Hazer & Bedell, 2006) relied entirely on undergraduate students instead of potential employers/HR personnel as participants. Furthermore, sample sizes in the aforementioned studies rarely exceeded 200 participants.

The goal of this study was therefore to close some of the gaps in the current literature and investigate the extent to which discrimination by HR personnel and hiring managers towards individuals with a former anxiety disorder is existent in the UK. We explored the following hypotheses:

> ***Hypothesis 1*:** Applicants with a disclosed former anxiety disorder will be viewed as less employable than people with hyperthyroidism or no disability, measured by the overall mean scores of all employability measurements.
> ***Hypothesis 2:*** Ratings of Resilience will mediate the effect of disability on anticipated competence.
> ***Hypothesis 3:*** Ratings of competence will mediate the effect of disability on recommendations.
> ***Hypothesis 4:*** Ratings of ambition will mediate the effect of disability on recommendations.

METHOD

Overview

We designed this study to examine the degree to which employment discrimination against people with anxiety disorders persists in the UK. We compared anticipated employability of potential employees shown as formerly having an anxiety disorder, hyperthyroidism, or no disability.

Participants

We recruited a total of 283 professionals with experience in recruiting and/or supervisory/managerial positions online through Prolific (https://www.prolific.co/). Comparable studies used Amazon's Mechanical Turk to recruit participants. However, Prolific provides subjects that are more naïve to experimental research designs, while ensuring a more ethnically diverse participant pool, higher transparency for participants and increased pre-screening options for researchers (Palan & Schitter, 2018; Peer et al., 2017). One hundred and seventy

participants are necessary to secure 80% Power for a two-way analysis of variance (ANOVA). This number was calculated prior to data collection using G*Power software (Version 3.1.9.6.; Faul et al., 2007) and assumed an effect size of F = .25, which is a medium effect size according to Cohen (1988). The final sample size was 257 after accounting for participants not completing the survey, and those who wanted their data removed after being debriefed. Participants were divided between genders (45.5% female, 54.5% male). The average participant was 38.22 years of age (SD = 10.93), ranging from 20 to 73 years and had spent 6.4 years in their current position, SD = 6.54, range 0-55. Most participants reported that they worked in middle management positions (42.8%), with junior management (24.5%), senior management (14.9%), administrative staff (5.8%), C-Level Executive (4.7), students (1.6%) and other (3.1%) comprising the remainder of the responses. The ethnic diversity of the sample was as follows: 78.2% were White British, 1.2% White Irish, 10.1% White Other, 1.2 % mixed Caribbean, .04 % mixed Other, 1.9 % Asian British Indian, 1.6 % Asian British Pakistani, .8 % Asian British and .8 % Asian British Other, 1.2 % Black African British, .8 % African Black Caribbean and .04 % counted themselves as other.

Study Design

Participants were randomly assigned to one of nine conditions in a 3 x 3 between-subjects design. There were 3 types of jobs and 3 types of illness to be evaluated: executive, manual, and administrative; anxiety disorder, hyperthyroidism, no illness. Participants were presented with a job posting for a specific position and asked to fill out a questionnaire evaluating the fitness of applicants for the specific role.

Procedure

The study was exclusively advertised on Prolific. Participants were asked to take part in a study investigating job selection processes, examining how opinions about applicants are constructed and how certain information is judged in ways that lead to consideration or exclusion of an individual from a larger pool of applicants. Prolific allows for pre-screening a custom sample. Inclusion criteria asked for UK residents aged 20 and above, who indicated having a leadership position or a position of power, supervisory duties or hiring experience. An upper age limit was not applied for screening. There are no studies involving recruiters which have indicated justification for, or have applied, an upper age limit. This was decided to be an ecologically valid decision as since 2011 there has been no default retirement age in the UK. Participants who decided to take part were redirected from Prolific,

through a link, directly to the survey, which was hosted at Qualtrics (https://www.qualtrics.com/uk). Participants were shown a participant information sheet and asked to sign a consent form, after which they were asked to provide demographic information. The first questions of the survey were rechecking the initial screening questions already applied on Prolific. We were unable to estimate a participant response rate. Prolific does not allow for such a function; furthermore, studies on Prolific are presented to all eligible participants, which means participants might see hundreds of studies per day. Calculating a response rate based on studies seen and having participated in, would be complex.

The executive job condition was represented by a job opening for the position of a managing director, the administrative job condition by a vacancy for a personal assistant. A vacancy for a barista was used for the manual condition. After participants read the job description, an applicants' cover letter and CV were presented. The CV contained a gap of six months in the section covering employment of the applicant. The cover letter contained the main experimental manipulation, indicating the disability of the applicant. In both the anxiety and hyperthyroidism condition, participants found the following sentence at the end of the cover letter:

> "You might be noticing a gap in my resume. I had to take some time off between my current full-time position and my previous position. I was given the diagnosis of an anxiety disorder/hyperthyroidism and had to take some time off to take care of this medical condition. Since that time, I have conquered my condition, am able to live a normal life and work at full time."

No statement of illness was made in the control cover letter. Hyperthyroidism was selected as the physical health issue, since it is marked by many symptoms overlapping with physical anxiety symptoms. As these two conditions are similar from an outside perspective, diminished evaluations could uneasily be justified by concerns of threads to productivity. Besides these variations, CVs were identical. Candidates' CV and cover letter were built in a way ensuring sufficient qualification for the vacant position, all had numerous previous full-time jobs, sufficient and even prestigious education. The CVs were made to look as if they were actual resumes that had only been anonymised, i.e., information regarding names of applicants, pictures, or addresses were missing, to exclude other possible discriminatory biases. After completion of the experiment, participants were debriefed regarding the actual purpose of this study and given the choice to reject their data from being included in analysis (this ethical necessity of the

study may have been a methodological limitation). Participants received 1.25 Great British Pounds (GBP) for participation in the survey. Participants were paid regardless of whether they chose to withdraw or keep their data. The full CVs and job adverts used in this study are available at doi:10.17605/OSF.IO/YJHKT.

Measures

Scales to evaluate candidates' anticipated employability were rooted in designs by Heilman and Okimoto (2008) and Kapoor (2017). Participants were asked to complete a survey of 10 items on which participants appraise the fictional job applicant. A full list of items can be found in table 1. Possible responses ranged on Likert-type scales from 1 (i.e., strongly disagree) to 9 (i.e., strongly agree). Questions of the scales were presented to participants in random order.

Table 1 Items and statements

Communication	"The candidate will communicate with co-workers and clients in a professional, concise and clear way."
Harmony	"The candidate will work in harmony with co-workers."
Ambition	"The candidate will be dedicated to the job, show eagerness to reach set goals and strive for achievement."
Competence	"The candidate is well suited for the job and will be able to handle the regular amount of daily stressors involved with this job."
Responsibility	"The candidate will take on responsibility and act in a responsible way."
Dependability	"The candidate is unlikely to take extra days off, or leave work early/arrive late."
Resilience	"The candidate will not easily crack under pressure or in a high stakes situation."
Independence	"The candidate will be able to work independently."
Recommendation	"I recommend this candidate for further consideration."

Elimination	"This candidate should NOT be excluded from further consideration."

Ethical Approval

Full approval was granted for this project by the King's College London Research Ethics Committee for Psychiatry, Nursing and Midwifery, Study Reference: HR-20/21-21856.

Data Set

Considering the importance of verifiable and dependable results, to allow for replication or secondary analyses of our findings, and in accordance with open science guidelines, all data supporting the results of this study are openly available on the OSF at doi:10.17605/OSF.IO/YJHKT

RESULTS

Shapiro-Wilk tests revealed no scale in no condition to be normally distributed (all $p < .001$) which violates one ANOVA assumption. However, ANOVAs are considered robust to violations of normality, especially when other assumptions are met (Basso et al., 2009; Blanca et al., 2017; Knief & Forstmeier, 2021; Lantz, 2012; Schmider et al., 2010). Three scales additionally violated Levene's test of equal variances, elimination ($F = 3.327$, $p = .002$), anticipated ambition ($F= 2.366$, $p = .018$) and anticipated competence ($F =2.654$, $p = .008$), which for these scales leaves two assumptions unmet.

Non-parametric alternatives to two-way ANOVAs exist but are weak in power and sensitive to differences in sample sizes, as well as not commonly used in psychological research, which allows for the alternative of running the standard two-way ANOVA, but interpreting the results with caution (Feys, 2016; Harrar et al., 2019; Liu et al., 2016; Marshall, n.d;). This is especially applicable given our large sample size, since, due to the central limit theorem, with increasing sample size, mean distribution will approximate normal distribution which increases the robustness of parametric tests (Fagerland, 2012; Fagerland & Sandvik, 2009; le Cessie et al., 2020).

Bayesian ANOVA were performed. Chosen were the overall scores, and the dependability measure, since it was the only measure from our explorative frequentist ANOVA yielding a significant main effect for the disability condition, which was the main target investigative subject of this study. Bayesian ANOVA followed procedures outlined by Rouder et al. (2012) and were conducted using JASP (Version 0.14.1) a free to use and open-source statistical and graphical software allowing for Bayesian hypothesis tests. The Bayesian ANOVA revealed no support for any of

the main effect models, or models including the interaction term. The Bayesian ANOVA revealed support for the main disability model over models including the interaction term and the null model.

Mediation analyses were conducted using the PROCESS package (Version 3.5; Hayes, 2020) for SPSS (IBM SPSS Statistics, Version 26). Regression-based bootstrap mediation analysis was performed as outlined by Hayes (2013), which yields the same reliability as SEM (Hayes & Scharkow 2013). Given the heterostadacity of the involved variables, robust standard errors of type HC3 (Davidson & McKinnon, 1993) were used for significance tests and confidence intervals.

When we included predicted dependability ratings as a mediator, disability significantly predicted the mediator (B = -.2625, p = .0072), which in turn predicted competence ratings significantly (B= .3270, p < .0001). The relationship between disability and predicted competence showed to be fully mediated by the dependability scores, as there was no longer a significant direct effect of the interaction on anticipated competency when we controlled for predicted dependability (B = -.0497, p =.5483), the indirect effect was also significant (ab = -.0858, 95%-CI [-.1627, -.0222]).

When we included predicted resilience ratings as a mediator, disability significantly predicted the mediator (B = -.2104, p = .0365) which in turn predicted competence ratings significantly (B= .3619, p < .0001). The relationship between disability and predicted competence showed to be fully mediated by the resilience scores as there was no longer a significant direct effect of the interaction on anticipated competency when we controlled for predicted resilience (B = -.0594, p = .4656). The indirect effect was also significant (ab = -.0761, 95%-CI [-.1586, -.0068]).

When we included anticipated competence as a mediator in the model, disability did not significantly predict the mediator (B = -.1355, p = .1054), which, however, predicted recommendation significantly (B = -.4609, p < .0001).

Some mediation experts advocate merely focusing on the indirect effect (Zhao et al., 2010; Rucker et al., 2011), but even this was not significant (ab = -.0625, 95% -CI [-.1386, .0097]). This speaks for the fact that ratings of predicted competence do not mediate recommendations.

When we included anticipated ambition as a mediator in the model, disability did not significantly predict the mediator (B = .0911, p = .2722) which, however, did predict recommendation significantly (B = .4956, p = .001).

The indirect effect was not significant (*ab* = -.0452, 95% -CI [-.0361, .1330]). This speaks for the fact that ratings of predicted ambition do not mediate recommendations either.

DISCUSSION

This study may be the first peer-reviewed paper to draw upon a UK sample of line managers and HR professionals to explore how discrimination surrounding people with a former anxiety disorder influences employability.

In detail, we examined the extent to which the label of an anxiety disorder led to less employability evaluations of employability, when compared to labels of hyperthyroidism or no disability. Interestingly, our results are in stark contrast to previous studies on the matter that used almost comparable or even identical designs and setups (Glozier, 1998; Hazer & Bedell, 2006; Kapoor, 2017; Koser et al., 1999) as we found neither the label of a former anxiety disorder nor physical health issue to influence overall employability of individuals. This speaks for the fact that, if employers are aware of an applicants' former mental or physical illness, they are just as likely to hire them as applicants without any such records. The hypothetical applicants were also equally likely to be recommended for further consideration and eliminated as their counterparts with a physical condition or no condition at all. The evaluations, which in theory, are thought to be based on objective measures of potential employees' training and qualifications, seem to be equal across the spectrum and not biased by any stigma. Our study used a large sample that yielded sufficient power, allowing for a widespread comparison of stigmatisation of mental illness across different job types, which revealed results in line with Kapoor (2017), finding no interaction effects between disability and job type. H1 was concerned with the overall mean scores. As this hypothesis had to be rejected, findings attest, for the UK labour market, absence of evidence of discrimination against people having formerly suffered from mental or physical disabilities. The underlying reason explaining this behaviour might be twofold. First, tides could be changing regarding public appraisal of mental health. The last study investigating mental illness and hiring discrimination in the UK was done by Glozier (1999) over two decades ago. Changed mindsets towards a more open assessment of mental health issues might be responsible for participants' equal ratings. Second, given the substantial prevalence of anxiety disorders, there is a fair chance that some participants either currently suffer themselves from one, did so in the past, or might know someone that has, which led them to reconsider their attitudes regarding these conditions and evaluate the hypothetical candidate correspondingly. Unexpectedly, effects of job type were

observed for certain measurement scales. Applicants' anticipated competence, ambition and recommendation for further consideration were evaluated more positively for manual positions. This could indicate that recruiters are more laissez-faire regarding manual positions but apply higher standards to administrative and executive positions. Additionally, we investigated the mediating role of some disability stereotypes in the formulations of competence expectations of potential candidates. This was inspired by findings of Heilman and Okimoto (2008), who ran a similar design with motherhood and employability. They found predicted competence to mediate effects of motherhood on recommendations, while ambition, they labelled it as "achievement striving" (p. 193), did not. Our results indicate nothing of the sort applying for disability. Competence ratings did not mediate the effect of disability type on recommendations; neither did ambition ratings. Findings from this mediation analyses yielded insights into the mechanisms underlying biased evaluations of applicants with a former anxiety disorder. Predicted deficits in both, dependability and resilience, fuelled the disability bias in recommendations. These findings suggest that especially traditionally biased perceptions of sufferers of mental disorders as being fragile, non-resilient and not dependable are some of the drivers behind the recommendation bias in employment contexts.

Importantly, the main part of this study yielded null results. Luckily, in this case, they are of heightened interest nonetheless, as the fact that employers seem not to discriminate could be interpreted as even more exciting than positive findings of discrimination. This is rare for null results. Positive findings appear to be of greater interest to publishers, researchers, and funders alike, which leads to a tendency in favour of positive results, a phenomenon known as the 'file-drawer problem', resulting in a substantial amount of research, which yielded negative results, remaining unpublished (American Psychological Association, 2021; Jacob et al., 2019; Munafo & Neil, 2016). Another reason behind the replication crisis, and important for discussing our findings, is the common reliance on null hypothesis significance testing (NHST) (Colling & Szűcs, 2021; Gelman, 2018; Szűcs & Ioannidis, 2017a). NHST compares data to the null hypothesis (H0) of no-effect, which is rejected in case the probability of observing the data or data more extremely deviating from H0 is below a certain threshold, usually a probability of 5%, assuming H0 is true. An alternative hypothesis (H1) is then favoured and H0 rejected. Otherwise, H0 is retained. However, this can only be interpreted as absence of evidence for H1, not as evidence in favour of H0. NHST has an array of drawbacks that have been described in detail by many previous studies on the debate, and by van Dongen and van Grooten (2021) summarised as "easy misuse and hidden

deficiencies" (p. 2). Deficiencies include overestimation of effect size (Gelman, 2018; Szűcs & Ionnadis, 2017b), neglecting of pre-data probabilities (Gigerenzer et al., 2008; Mertens & Recker, 2020), unsuitability for large data sets (Hofmann, 2015; Nickerson, 2000), low power (Button et al., 2013; Dumas-Mallet et al., 2017; Gelman & Carlin, 2014) and the argument that null hypotheses are always false, since, with enough power, the null hypothesis can always be rejected (Kirk, 1996). Often cited are also misunderstandings of fundamentals behind *p*-values, such as the conditional probability that H0 is correct when data show extreme values of the test measure, when actually the *p*-value measures conditional probability that the test statistic takes on more extreme values than the observed value when H0 is correct (Haller & Kraus, 2002). Misuse of NHST include malicious scientific practices associated with it, such as post-hoc storytelling, also known as hypothesis testing after results are known (HARKing) (Chambers et al., 2014; Cockburn et al. 2018; Mayer, 2019) and *p*-value hacking (Crane, 2018; Ioannidis, 2019; Lyu et al., 2018). Furthermore, NHST encourages an all or nothing approach interpreting the H0 or H1 (Szűcs & Ioannidis, 2017a). Alternatives are considerations of raw data, Fiducial interference, and Bayesian approaches (Colling & Szűcs, 2021; Hannig et al., 2016; Quintana & Williams, 2018; Wagenmakers et al., 2017). Bayesian statistics are believed to outperform traditional frequentist approaches, as they allow for parameter estimations and hypothesis descriptions. Furthermore, prior information can be factored into the model, which is of interest when considering past study results (Wagenmakers et al., 2017). The Bayesian approach offers quantification of the likelihood of the data under the H0 in comparison to H1, considering a prior probability, while also including parameter estimation, allowing to address effect size (Dienes & Mclatchie, 2018). Szűcs and Ioannidis (2017a) also propose, besides applying Bayesian approaches, making data, power calculations and sample sizes publicly available and publishing results regardless of statistical significance, all of which we complied to.

Limitations

Even though our findings do not suggest discrimination in hiring procedures, they should be enjoyed with caution and may not translate to the real world. The present research was conducted in an analogue manner, with participants making decisions regarding fictitious people, which, even though presented as being real applicants, were still assessed in the frame of a research project. This is fundamentally different than appraising potential candidates to work for one's own firm. Furthermore, the participants were only given marginal information, i.e., a CV and a

cover letter, therefore other stages of the hiring process such as interviews and further social interaction, were excluded in this experimental design. Additionally, participants were each only presented with a single candidate, not allowing for direct comparison. Socially desired answers might furthermore have influenced participants responses in this study, potentially rivalling with their real-world behaviour. Furthermore, our study only used a six-month resume gap – longer or shorter periods of absence from the labour market might be interpreted differently.

Conclusion

Our study suggests little to no evidence for discrimination against anxiety disorders in some areas of job selection. Applicants with a former mental illness were evaluated to be just as employable, were recommended and rejected just as often as their counterparts with the same qualification, having hyperthyroidism or no disability.

REFERENCES

Ameri, M., Schur, L., Adya, M., Bentley, F. S., McKay, P., & Kruse, D. (2017). The disability employment puzzle: A field experiment on employer hiring behavior. *ILR Review*, *71*(2), 329–364. https://doi.org/10.1177/0019793917717474

American Psychological Association. (2021). file-drawer problem. Retrieved from: https://dictionary.apa.org/file-drawer-problem. Accessed on July 7th, 2021.

Baert, S. (2016). Wage subsidies and hiring chances for the disabled: Some causal evidence. *European Journal of Health Economics, 17*, 71–86.

Baert, S. (2018). Hiring discrimination: An overview of (almost) all Correspondence experiments since 2005. *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, 63–77. https://doi.org/10.1007/978-3-319-71153-9_3

Baert, S., De Visschere, S., Schoors, K., Vandenberghe, D., & Omey, E. (2016). First depressed, then discriminated against? Social Science & Medicine, 170, 247–254.

Basso, D., Pesarin, F., Salmaso, L., & Solari, A. (2009). Synchronized permutation tests in two-way ANOVA. *Permutation Tests for Stochastic Ordering and ANOVA*, 133–172. https://doi.org/10.1007/978-0-387-85956-9_6

Baxter, A. J., Scott, K. M., Vos, T., & Whiteford, H. A. (2013). Global prevalence of anxiety disorders: a systematic review and meta-regression. *Psychological medicine*, *43*(5), 897-910

Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., & Whiteford, H. A. (2014). The global burden of anxiety disorders in 2010. *Psychological Medicine*, *44*(11), 2363–2374. https://doi.org/10.1017/s0033291713003243

Bell, B. S., & Klein, K. J. (2001). Effects of disability, gender, and job level on ratings of job applicants. *Rehabilitation Psychology*, *46*(3), 229.

Bjørnshagen, V. (2021). The mark of mental health problems. A field experiment on hiring discrimination before and during COVID-19. *Social Science & Medicine*, *283*, 114181.

Blanca, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 2017, vol. 29, num. 4, p. 552-557*.

Bricout, J. C., & Bentley, K. J. (2000). Disability status and perceptions of employability by employers. *Social Work Research*, *24*(2), 87-95

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Confidence and precision increase with high statistical power. *Nature Reviews Neuroscience*, *14*(8), 585–585. https://doi.org/10.1038/nrn3475-c4

Cohen, J (1988). *Statistical Power Analysis for the Social Sciences (2nd. Edition)*. Hillsdale, New Jersey, Lawrence Erlbaum Associates.

Chambers, C., Feredoes, E., D. Muthukumaraswamy, S., & J. Etchells, P. (2014). Instead of "playing the game" it is time to change the rules: Registered reports at AIMS neuroscienc and beyond. *AIMS Neuroscience*, *1*(1), 4–17. https://doi.org/10.3934/neuroscience.2014.1.4

Cockburn, A., Gutwin, C., & Dix, A. (2018). Hark no more. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3173574.3173715

Colling, L. J., & Szűcs, D. (2021). Statistical inference and the replication crisis. *Review of Philosophy and Psychology*, *12*(1), 121–147. https://doi.org/10.1007/s13164-018-0421-4

Crane, H. (2018). The impact of p-hacking on "redefine statistical significance." *Basic and Applied Social Psychology*, *40*(4), 219–235. https://doi.org/10.1080/01973533.2018.1474111

Dalgin, R. S. & Bellini, J. (2005). *Disability disclosure in an employment interview: Impact on employers' hiring decisions and views of employability*. Syracuse University.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Demming, C. L., Jahn, S., & Boztug, Y. (2017). Conducting mediation analysis in marketing research. *Marketing ZFP*, *39*(3), 76–98. https://doi.org/ 10.15358/0344-1369-2017-3-76

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, *25*(1), 207–218. https://doi.org/10.3758/s13423-017-1266-z

Dumas-Mallet, E., Button, K. S., Boraud, T., Gonon, F., & Munafò, M. R. (2017). Low statistical power in Biomedical science: A review of three human RESEARCH DOMAINS. *Royal Society Open Science*, *4*(2), 160254. https://doi.org/10.1098/rsos.160254

Essau, C. A., Lewinsohn, P. M., Lim, J. X., Ho, M.-ho R., & Rohde, P. (2018). Incidence, recurrence and comorbidity of anxiety disorders in four major developmental stages. *Journal of Affective Disorders*, *228*, 248–253. https://doi.org/10.1016/j.jad.2017.12.014

Fagerland, M. W. (2012). t-tests, non-parametric tests, and large studies—a paradox of statistical practice?.*BMC Medical Research Methodology*, *12*(1), 1-7.

Fagerland, M. W., & Sandvik, L. (2009). Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemporary clinical trials*, *30*(5), 490-496.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power, Version 3.1.9.6., Düsseldorf, Germany. Retrieved from https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

Feria, L. J., End, C. M., & Yadlosky, L. B. (2014). Depression may bring rejection even in volunteer selection situations. *North American Journal of Psychology*, *16*(2).

Feys, J. (2016). Nonparametric tests for the interaction in two-way factorial designs using R. *The R Journal*, *8*(1), 367-378.

Gelman, A. (2018). The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Personality and Social Psychology Bulletin*, *44*(1), 16–23. https://doi.org/10.1177/0146167217729162

Gelman, A., & Carlin, J. (2014). Beyond power calculations. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Glozier, N. (1998). Workplace effects of the stigmatization of depression. *Journal of Occupational & Environmental Medicine*, *40*(9), 793–800. https://doi.org/10.1097/00043764-199809000-00008

Gouvier, W. D., Sytsma-Jordan, S., & Mayville, S. (2003). Patterns of discrimination in hiring job applicants with disabilities: The role of disability type, job complexity, and public contact. *Rehabilitation psychology*, *48*(3), 175.

Haller, H., & Kraus, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7*(1), 1–20.

Hannig, J., Iyer, H., Lai, R. C., & Lee, T. C. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, *111*(515), 1346-1361.

Harrar, S. W., Ronchi, F., & Salmaso, L. (2019). A comparison of recent nonparametric methods for testing effects in Two-by-two factorial designs. *Journal of Applied Statistics*, *46*(9), 1649–1670. https://doi.org/10.1080/02664763.2018.1555575

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. The Guilford Press.

Hayes, A. F. (2020). PROCESS macro for SPSS, Version 3.5., Alberta, Canada. Retrieved from https://www.processmacro.org/download.html

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*(10), 1918–1927.

Hazer, J. T., & Bedell, K. V. (2000). Effects of seeking accommodation and disability on preemployment evaluations. *Journal of Applied Social Psychology*, *30*(6), 1201-1223.

Heilman, M. E., & Okimoto, T. G. (2008). Motherhood: A potential source of bias in employment decisions. *Journal of Applied Psychology*, *93*(1), 189–198. https://doi.org/10.1037/0021-9010.93.1.189

Hipes, C., Lucas, J., Phelan, J. C., & White, R. C. (2016). The stigma of mental illness in the labor market. *Social Science Research*, *56*, 16–25. https://doi.org/10.1016/j.ssresearch.2015.12.001

Hofmann, M. A. (2015). Searching for effects in big data: Why p-values are not advised and what to use instead. *2015 Winter Simulation Conference (WSC)*. https://doi.org/10.1109/wsc.2015.7408210

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65-70.

Iacobucci , D., Saldanha, N., & Deng , X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology*, *17*(2), 139–153. https://doi.org/10.1016/s1057-7408(07)70020-7

IBM Corp (2019). IBM SPSS Statistics for MacOS, Version 26.0. Armonk, NY.

In praise of replication studies and null results. (2020). *Nature*. *578*, 489-490. *https://doi.org/10.1038/d41586-020-00530-6*

Ioannidis, J. P. (2019). What have we (not) learnt from millions of scientific papers with p values? *The American Statistician*, *73*(sup1), 20–25. https://doi.org/10.1080/00031305.2018.1447512

Jacob, R. T., Doolittle, F., Kemple, J., & Somers, M.-A. (2019). A framework for learning from null results. *Educational Researcher*, *48*(9), 580–589. https://doi.org/10.3102/0013189x19891955

JASP Team (2020). JASP, Version 0.14.1. Amsterdam, The Netherlands. Retrieved from: https://jasp-stats.org/download/

Kapoor, A. (2017). Depressed people need not apply: Mental health stigma decreases perceptions of employability of applicants with depression. *Yale Review of Undergraduate Research in Psychology,* 84-94. https://cpb-us-w2.wpmucdn.com/campuspress.yale.edu/dist/a/1215/files/2017/06/Arunima-1amuxqj.pdf

Kessler, R. C., & Üstün, T. B. (2004). The World Mental Health (WMH) survey INITIATIVE version of the World Health Organization (WHO) composite international Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, *13*(2), 93–121. https://doi.org/10.1002/mpr.168

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593-602.

Kessler, R., Angermeyer, M., Anthony, J. C., Graaf, R. D. E., Demyttenaere, K., & Gasquet, I. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, *6*(3), 168–76.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746–759. https://doi.org/10.1177/0013164496056005002

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01587-5

Koser, D. A., Matsuyama, M., & Kopelman, R. E. (1999). Comparison of a physical and a mental disability in employee selection: an experimental examination of direct and moderated effects. *North American Journal of Psychology*, *1*(2), 213-222

Lantz, B. (2012). The impact of sample non-normality on ANOVA and alternative methods. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 224–244. https://doi.org/10.1111/j.2044-8317.2012.02047.x

le Cessie, S., Goeman, J. J., & Dekkers, O. M. (2020). Who is afraid of non-normal data? Choosing between parametric and non-parametric tests. *European journal of endocrinology*, *182*(2), E1-E3.

Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test?. *Korean Journal of Anesthesiology*, *71*(5), 353.

Liu, X., Guo, J., Zhou, B., & Zhang, J.-T. (2016). Two simple tests for Heteroscedastic TWO-WAY ANOVA. *Statistics Research Letters*, *5*, 6. https://doi.org/10.14355/srl.2016.05.002

Lyu, Z., Peng, K., & Hu, C.-P. (2018). P-Value, confidence intervals, and statistical Inference: A new dataset of misinterpretation. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00868

Marshall, E. (n.d.). Two-Way ANOVA in SPSS, Statstutor Community Project., Department of Engineering and Maths, Sheffield Hallam University, https://maths.shu.ac.uk/mathshelp/Stats%20support%20resources/Resources/Testing%20means/Comparing%20groups/ANOVA/SPSS/stcp-marshall-ANOVA2.pdf

Mayer, B. (2019). Hacking, picking, harking and co.: Lessons from the replication crisis in psychology. In *Retreat of the Interfaculty Research Cooperation (IRC) "Decoding Sleep"*. Bern, Switzerland; Universität Bern.

Mertens, W., & Recker, J. (2020). New guidelines for null hypothesis smeignificance testing in hypothetico-deductive is research. *Journal of the Association for Information Systems*, *21*, 1072–1102. https://doi.org/10.17705/1jais.00629

Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., & Ustun, B. (2007). Depression, chronic diseases, and decrements in HEALTH: Results from the World Health Surveys. *The Lancet*, *370*(9590), 851–858. https://doi.org/10.1016/s0140-6736(07)61415-9

Munafò, M., & Neill, J. (2016). Null is beautiful: On the importance of publishing null results. *Journal of Psychopharmacology*, *30*(7), 585–585. https://doi.org/10.1177/0269881116638813

National Health Service. (2018). *Signs of an Anxiety Disorder*. NHS choices. Retrieved from: https://www.nhs.uk/mental-health/feelings-symptoms-behaviours/feelings-and-symptoms/anxiety-disorder-signs/.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241–301. https://doi.org/10.1037/1082-989x.5.2.241

Organisation for Economic Co-Operation and Development. (2012). (rep.). *Sick on the Job? Myths and Realities about Mental Health and Work*. OECD.

Ormel, J., Petukhova, M., Chatterji, S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Bromet, E. J., Burger, H., Demyttenaere, K., de Girolamo, G., Haro, J. M., Hwang, I., Karam, E., Kawakami, N., Lépine, J. P., Medina-Mora, M. E., Posada-Villa, J., Sampson, N., Scott, K., … Kessler, R. C. (2008). Disability and treatment of specific mental and physical disorders across the world. *British Journal of Psychiatry*, *192*(5), 368–375. https://doi.org/10.1192/bjp.bp.107.039107

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27

Pearson, V., Ip, F., Hui, H., & Yip, N. (2003). To tell or not to tell; disability disclosure and job application outcomes. *Journal of Rehabilitation*, *69*(4), 35.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153-163

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, *56*(3), 345–365. https://doi.org/10.1111/jcpp.12381

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide USING JASP. *BMC Psychiatry*, *18*(1). https://doi.org/10.1186/s12888-018-1761-4

Rewarding negative results keeps science on track. (2017). *Nature* **551,** 414 (2017). https://doi.org/10.1038/d41586-017-07325-2

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356-374.

RStudio Team (2020). RStudio: Integrated development for R. RStudio, Version 1.4.1106. Boston, MA. Retrieved from: http://www.rstudio.com/.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology*, *6*(4), 147–151. https://doi.org/10.1027/1614-2241/a000016

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, *43*(2), 476-493. https://doi.org/10.1093/ije/dyu038

Stein, D., Scott, K., de Jonge, P., & Kessler, R. (2017). Epidemiology of anxiety disorders: From surveys to nosology and back. *Dialogues Clin Neuroscience*, *19*(2), 127–136. https://doi.org/10.31887/dcns.2017.19.2/dstein

Sterkens, P., Baert, S., Rooman, C., & Derous, E. (2021). As if it weren't hard enough already: Breaking down hiring discrimination following burnout. *Economics & Human Biology*, 101050.

Szucs, D., & Ioannidis, J. P. (2017a). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00390

Szucs, D., & Ioannidis, J. P. (2017b). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3). https://doi.org/10.1371/journal.pbio.2000797

Trades Union Congress. (2017). (rep.). *Mental health and employment*. Retrieved from https://www.tuc.org.uk/sites/default/files/Mental_Health_and_Employment.pdf

van Aert, R. C., Wicherts, J. M., & van Assen, M. A. (2019). Publication bias in Meta-analyses from psychology and medicine: A Meta-meta-analysis. https://doi.org/10.31222/osf.io/czdp3

van Dongen, N. N. D., & van Grootel, L. (2021). Overview on the null hypothesis significance test. https://doi.org/10.31234/osf.io/hwk4n

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wilson, B. M., Harris, C. R., & Wixted, J. T. (2020). Science is not a signal detection problem. *Proceedings of the National Academy of Sciences*, *117*(11), 5559–5567. https://doi.org/10.1073/pnas.1914237117

World Health Organization. (2020). *International classification of diseases for mortality and morbidity statistics* (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/en

Zissi, A., Rontos, C., Papageorgiou, D., Pierrakou, C., & Chtouris, S. (2007). Greek employers' attitudes to employing people with disabilities: Effects of type of disability. *Scandinavian Journal of Disability Research*, *9*(1), 14-25

Data availability statement: The data that support the findings of this study are openly available on the Open Science Framework (OSF) at doi:10.17605/OSF.IO/YJHKT

Acknowledgements: Thank you to Myanna Duncan and James Findon for reviewing an early draft of this paper. We would also like to thank the two anonymous reviewers for their constructive comments and feedback. Funding for payment of participants was provided through King's College London bench fees, supporting project costs directly and exclusively. There are no additional funding sources or conflicts of interest to report.