Article:

Firth, N. orcid.org/0000-0003-1984-6869, Barkham, M. orcid.org/0000-0003-1687-6376, Delgadillo, J. orcid.org/0000-0001-5349-230X et al. (2 more authors) (2023) The role of socioeconomic deprivation in explaining neighborhood and clinic effects in the effectiveness of psychological interventions. Journal of Consulting and Clinical Psychology, 91 (2). pp. 82-94. ISSN 0022-006X

https://doi.org/10.1037/ccp0000784

Abstract

Objective: Treatment outcomes are known to vary according to therapist and clinic/organization (therapist effect, clinic effect). Outcomes may also vary according to the neighborhood where a person lives (neighborhood effect), but this has not previously been formally quantified. Evidence suggests that deprivation may contribute to explaining such cluster effects. This study aimed to 1) simultaneously quantify neighborhood, clinic, and therapist effects on intervention effectiveness, and 2) determine the extent to which deprivation variables explain neighborhood and clinic effects. **Method:** The study used a retrospective, observational cohort design, with a high intensity psychological intervention sample (*N* = 617,375), and a low intensity psychological intervention sample (*N* = 773,675). Samples each included 55 clinics, 9,000-10,000 therapists/practitioners and over 18,000 neighborhoods in England. Outcomes were post-intervention depression and anxiety scores, and clinical recovery. Deprivation variables included individual employment status, domains of neighborhood deprivation, and clinic-level mean deprivation. Data were analyzed using cross-classified multilevel models. **Results:** Unadjusted neighborhood effects of 1-2% and unadjusted clinic effects of 2-5% were detected, with proportionally larger effects for low intensity interventions. After controlling for predictors, adjusted neighborhood effects of 0.0-0.1% and clinic effects of 1-2% remained. Deprivation variables explained a significant proportion of the neighborhood effect (80-90% of neighborhood variance), but not clinic effect. The majority of neighborhood variance could only be explained by a shared effect of baseline severity and socioeconomic deprivation variables. **Conclusions:** People in different neighborhoods respond differently to psychological intervention, and this clustering effect was mainly explained by socioeconomic factors. People also respond differently according to the clinic they access, but this could not be completely explained by deprivation in the current study.

Keywords: deprivation; neighborhood effect; therapist effect; organization effect; IAPT

Public Health Significance Statement

This study demonstrates that the clinical effectiveness of psychological treatment systematically varies according to the neighborhood where a patient lives, as well as the clinic that the patient accesses. Results indicate that socioeconomic deprivation is key in understanding these cluster effects, particularly between neighborhoods. It is important that clinical effectiveness targets are adjusted for socio-clinical context, whilst also working to improve treatment effectiveness for those patients currently experiencing poorer outcome.

Introduction

It is well evidenced that some patients find psychological treatment more helpful than others, and that certain patient characteristics are predictive of better or poorer outcomes (Constantino et al., 2021). A robust and growing body of research also indicates that after controlling for patient differences, some therapists have significantly better outcomes than others on average, and the same is true for different healthcare clinics/organizations (Falkenström et al., 2018; Firth et al., 2019; Johns et al., 2019; Wampold & Owen, 2021). These types of clustering have been termed therapist effects and clinic or organization effects, although they do not necessarily imply direct causative effects. Therapist effects typically account for around 5-10% of variance in outcomes (Baldwin & Imel, 2013; Johns et al., 2019), with emerging evidence finding similar or slightly smaller estimates of clinic effects (Falkenström et al., 2018; Firth et al., 2019).

Given the complex context of psychological interventions, there may be other clustering effects on outcomes, such as the neighborhood where an individual lives. Neighborhood effects have been demonstrated in prevalence studies of mental and physical illness (Freedman & Woods, 2013; Richardson et al., 2015; Silva et al., 2016). Conversely, although some neighborhood variables have been investigated in psychological treatment research, no studies to date have quantified neighborhood effects as a proportion of unexplained outcome variance, in line with the now-standard formal therapist effect definition (Wampold & Owen, 2021). Clinic effects may be confounded by unmodeled neighborhood effects, given clinics often cover a distinct geographic area (Firth et al., 2019). Multilevel analyses are required to ensure that each clustering *level* is accounted for appropriately. To our knowledge, there have been no studies that have simultaneously analyzed patient, therapist, neighborhood, and clinic level clustering of outcomes.

Where clustering is detected, it is important to understand which variables explain this clustering. Socioeconomic deprivation is a good candidate for explaining variability at the neighborhood level in particular; it has typically been one of the main explanatory factors in studies of neighborhood effects on mental/physical illness prevalence (O'Donoghue et al., 2016; Pickett & Pearl, 2001; Silva et al., 2016; Wilson, 1987). Deprivation has also been associated with poorer access to mental health treatment, and poorer psychological treatment outcomes (Delgadillo et al., 2018; Finegan et al., 2018, 2020; Muntaner et al., 2004; Poole et al., 2014). It is conceivable that deprivation may in part explain clinic effects (e.g., limited budgets for clinics in deprived regions could lead to staff shortages, lack of appropriate treatment locations/facilities, etc.), although research to date has focused more on organizational climate and culture as explanatory factors (Brand et al., 2012; Falkenström et al., 2018). We expected that the clinic effect would be most closely associated with organizational factors (Brand et al., 2012; Falkenström et al., 2018), and therefore that deprivation variables in this study would likely have a weaker association with clinic effects, compared with neighborhood effects.

Socioeconomic deprivation may be defined as a lack of social or material conditions or resources important to facilitate a good quality of life, and is closely linked with the concept of poverty (Poverty and Social Exclusion UK, 2016; Townsend, 1979, 1987). Deprivation is multifaceted (American Psychological Association Task Force on Socioeconomic Status, 2007; Ministry of Housing, Communities & Local Government, 2019) and may act at different levels, such as the individual (e.g., individual education, occupation), household (e.g., household income, living conditions), neighborhood (e.g., neighborhood level of crime, transport links), clinical organization (e.g., clinic funding and resources), large area jurisdiction (e.g., wealth and resources of local or national health authorities or governing bodies), etc. However, previous studies have not typically controlled for different types of deprivation or for clustering effects, increasing the risk of confounding and inaccurate inference.

Effects of deprivation are important as they may maintain or exacerbate inequality (and even mortality; Longevity Science Panel, 2018). Indeed, health care systems may themselves further increase inequality, for example via intervention generated inequalities and the inverse care law. This is particularly relevant for care contexts that use outcomes-based-payment models (NHS England & NHS Improvement, 2017). In theory, clinics that are struggling to meet outcome targets due to working in areas of deprivation may then receive less funding, potentially exacerbating an already vicious cycle (Webb & Bywaters, 2018).

The current study aimed to investigate to what extent psychological treatment outcomes are clustered at the neighborhood and clinic levels (neighborhood and clinic effects), before and after controlling for explanatory and control variables at each level. The study also aimed to assess the extent to which the neighborhood and clinic effects could be explained by variables reflecting aspects of socioeconomic deprivation. These aims were separately investigated for two types of intervention (high intensity and low intensity; see Methods).

We hypothesized that patient outcomes would be clustered according to both the neighborhood where they lived, and the clinic where they received treatment (in addition to expected clustering according to therapist). We also hypothesized that neighborhood clustering would significantly reduce after controlling for deprivation related variables (indicating an explanatory effect), but clinic level clustering would not significantly change.

Methods

Design, Transparency, and Openness

A retrospective observational cohort design was used to analyze routine clinical data. The study was approved by the Health Research Authority (reference 19/HRA/6096). Ethical approval by a National Health Service Research Ethics Committee was not required given the study used only anonymized retrospective routine care data. The [ANONYMISED INSTITUTION] Research Ethics Committee approved a self-declaration to this effect (reference [ANONYMISED]). The study design, hypotheses, and analysis plan were pre-registered ([ANONYMISED FOR PEER REVIEW]). Clinical data were obtained on condition of a data sharing agreement that does not permit public availability of data. Analysis code is not applicable (analyses were performed using user interfaces), but full model specifications are shown in Supplementary Material.

Clinical Context

The current study focused on patients receiving treatment from the Improving Access to Psychological Therapies (IAPT) national program in England. IAPT aims to deliver evidence-based manualized psychological interventions for common mental health disorders, using a stepped care framework (Clark, 2011, 2018). This includes offering *low intensity* (LI) interventions such as guided self-help, bibliotherapy, and computerized mental health interventions, as well as *high intensity* (HI) interventions such as cognitive behavioral therapy, interpersonal and psychodynamic psychotherapies, and person-centered-experiential counseling for depression. HI interventions are delivered by psychological therapists, counselors, and clinical psychologists, and LI interventions are delivered by psychological wellbeing practitioners (PWPs). Although the role of PWPs is akin to that of coaches with specific training in brief interventions, for simplicity the word “therapist” is used in this study to refer to both qualified HI therapists and LI PWPs. Research has previously demonstrated therapist effects for PWP delivered interventions of comparable size to those detected for qualified therapists (Firth et al., 2015; Green et al., 2014).

All IAPT clinics are required to meet the same standard national recovery target. As the national program for delivering frontline psychological interventions, IAPT has a large number of clinics situated across the country, using the same treatment framework across clinics. In some areas, patients have the choice of more than one IAPT clinic, further enabling differentiation between neighborhood and clinic effects. This manualized treatment provision context is also very suitable for comparison of effects between LI and HI interventions.

Sample and Participants

IAPT clinics in England use data management software systems to record clinical activity and outcomes. The research team approached a total of 65 IAPT clinics that used a specific data management system (total population sampling). Of these, 55 clinics consented to provide anonymized data. Data were not limited by timeframe, to maximize sample and cluster sizes. Data covered an approximate 12-year range (end of 2008 to start of 2020; IAPT was implemented nationally in 2008).

Participants were patients who had been discharged after receiving psychological treatment in IAPT clinics. Inclusion criteria required that patients had attended at least two sessions (defined by IAPT as accessing treatment – i.e., an initial assessment session plus at least one treatment session), that the intervention included only individual treatment sessions (excluding group or couple treatment), and that there were valid data for the dependent variable and all independent variables (see Measures). For the purposes of this study, an intervention was defined as a series of two or more consecutive sessions at a certain step of treatment (either LI or HI). We used session-by-session data with step identifiers to delineate LI and HI interventions. Only the first eligible intervention per patient was included, to fulfil the statistical assumption of independent observations for each patient. Figure 1 shows the inclusion process. Of 971,765 HI interventions and 1,058,230 LI interventions with at least two attended sessions, there were 617,375 HI interventions (64%) and 773,675 LI interventions (73%) included in the final LI and HI samples. A smaller proportion of HI interventions were selected in the final sample because patients were more likely to have had multiple eligible HI interventions (and only the first for each patient was selected in the final sample). Because patients may have attended interventions at both LI and HI, a given patient may be represented in both samples (in other words, the LI and HI samples are likely to overlap with regards to patients represented, although not with regards to specific treatment interventions or sessions). All patients were age 16 or over.

The HI sample was clustered within 18,701 neighborhoods. HI patients were seen by 10,254 therapists, across 55 clinics. There was a mean of 33 patients per neighborhood, a mean of 60 patients per therapist, and a mean of 186 therapists per clinic. The LI sample was clustered within 18,974 neighborhoods. Because patients receiving LI interventions sometimes see more than one PWP, a "primary PWP" was defined as the PWP who delivered the highest proportion of attended sessions. LI patients were seen by 9,093 primary PWPs, across 55 clinics. There was a mean of 41 patients per neighborhood, a mean of 85 patients per primary PWP, and a mean of 165 primary PWPs per clinic. The sample thus conformed to sample size recommendations of at least 1000-1500 patients, 50 top level units, and 30-50 lower level units per top-level cluster (Hox, 2010; Moineddin et al., 2007; Schiefele et al., 2017).

The sample was selected using list-wise deletion. Relative benefits and costs of list-wise deletion versus approaches such as multiple imputation vary as a function of patterns of missingness and power (Allison, 2001; Pepinsky, 2018). In this study, the major limitation on sample size was to include only one episode per patient, to preserve independence of observations. Because of this, multiple imputation would have increased the sample only marginally, and power was considered sufficient with complete cases. There was minimal outcome missingness, meaning listwise deletion should be relatively unbiased if data are missing at random (MAR) or missing completely at random, whilst multiple imputation can be more biased, less efficient, and have worse coverage when data are missing not at random (MNAR; Allison, 2001; Pepinsky, 2018). It is not possible to directly test MNAR vs MAR assumptions, but MNAR was considered possible. In combination with other pragmatic considerations, list-wise deletion was chosen.

Measures

All IAPT clinics are required to collect and report a broad set of standard measures known as the Minimum Dataset (MDS). Measures from the MDS were supplemented with publicly available neighborhood data from the national 2011 Census (Office for National Statistics, n.d.), described below.

Dependent variables

The primary outcome variable in this study was post-treatment depression symptoms measured using the Patient Health Questionnaire (PHQ-9), a validated 9-item depression measure scored from 0-27, with higher scores indicating greater severity (Cameron et al., 2008; Kroenke et al., 2001; Martin et al., 2006). As a secondary outcome, anxiety scores were measured using the Generalized Anxiety Disorder-7 (GAD-7), a validated 7-item anxiety measure scored from 0-21, with higher scores indicating greater severity (Spitzer et al., 2006). Patients attending IAPT clinics are expected to complete both measures at each appointment. IAPT uses cut-offs of ≥10 for the PHQ-9 and ≥8 for the GAD-7 to detect clinically significant symptoms.

IAPT defines patient recovery as occurring where a patient scoring above clinical threshold for depression and/or anxiety symptoms finishes treatment with scores below the threshold for both depression and anxiety symptoms (The National Collaborating Centre for Mental Health, 2021). Thus, a secondary analysis using this recovery definition was conducted, only included patients above threshold pre-treatment (high intensity = 541,627, low intensity = 688,940). Recovery was a binary outcome (recovered vs. not recovered). Finally, a post-hoc supplementary analysis was conducted with reliable and clinically significant change (RCSI) as the outcome and reliable change indices of ≥ 6 for the PHQ-9 and ≥ 4 for the GAD-7 (Jacobson & Truax, 1991).

Explanatory variables

Explanatory independent deprivation variables were as follows. Patient level data on employment status (*employed, retired, other-role* [*student, volunteer,* or *homemaker*]*, unemployed-seeking, unemployed-benefits,* and *unemployed-no-benefits*) were included as part of the minimum dataset. Neighborhood level data were also included. A neighborhood is a small local area - in this study neighborhood is defined by the UK government definition of a Lower Layer Super Output Area (LSOA), approximating 1500 people or 650 households (NHS Digital, 2021). LSOA census data were linked to clinical data using patients' LSOA.

Neighborhood-level deprivation variables were the seven domains of the 2019 Indices of Deprivation (IMD; Ministry of Housing, Communities & Local Government, 2019) (but not the overall Index of Multiple Deprivation score). These domains are as follows: Income; Employment; Health Deprivation and Disability; Education, Skills, and Training; Barriers to Housing and Services; Crime; Living Environment. Variables were specified as a relative percentage, in order to minimize parameters whilst retaining data granularity and transferability to other metrics (e.g., deciles, quintiles). In other words, an IMD-Income score of 2% indicates that the neighborhood is in the top 2% of neighborhoods for income deprivation (thus lower percentages indicate greater deprivation).

At the clinic level, clinic mean deprivation was calculated, intending to reflect the average deprivation experienced by the population as a whole served by the clinic. This was designed to capture any potential aggregate effect of deprivation at the clinic level that may differ from effects at the individual or neighborhood level (e.g., Firth et al., 2019; Saxon & Barkham, 2012), and that may not be limited only to the direct effects of patients included in the current sample. This was a proxy measure calculated by taking the Index of Multiple Deprivation composite score for each patient (Ministry of Housing, Communities & Local Government, 2019) and averaging across all patients in the full dataset who had been referred to that clinic. There was a >.99 correlation between this measure and an equivalent measure for the included sample only (see Supplementary Material).

Control variables

Continuous control variables at the patient level from the minimum dataset included baseline PHQ-9, GAD-7 and Work and Social Adjustment Scale (WSAS) scores, patients' age, percentage of sessions delivered face-to-face, and treatment start date. The WSAS is a validated 8-item measure of functional impairment, scored from 0-40 with higher scores indicating greater severity (Mundt et al., 2002). Categorical measures of, ethnicity, disability, and patient-reported gender (male, female) were also included. Ethnicity included six categories, corresponding to Office for National Statistics (2016) categories (*White, Mixed/Multiple, Asian, Black, Other*), plus a *not reported* category (the latter comprising 9% of sample). Preliminary analysis found that this operationalization demonstrated superior model fit in compared with dichotomous alternatives (White British/Other or White/Other). Disability included three categories: *has disability, no perceived disability,* and *not reported* (22% of sample). In the high intensity treatment sample, a variable was included for whether patients had already accessed 2 or more sessions of low intensity treatment previously within the same referral (yes, no).

There were no reliable therapist-level variables available in the dataset. However, modeling a therapist level accounts for all variance ostensibly associated with the therapist, even if it is caused by unmeasured therapist attributes.

At the neighborhood level, census data on White ethnic density (W-ED) was included as a control variable. This was measured as the percentage of people in a neighborhood who identify as

White.

At the clinic level, clinic W-ED was included as a control variable. Clinic W-ED was the percentage of patients accessing the clinic who identify as White. At both neighbourhood and clinic levels, preliminary analysis found superior or equivalent model fit for the W-ED operationalization compared with an alternative (White British ethnic density).

For this study, deprivation variables were considered to be employment status, the seven neighborhood deprivation domains, and clinic mean IMD. Although ethnicity and disability correlate with deprivation in some contexts, these were not considered deprivation variables in this study. Clinic level variables (clinic W-ED and clinic mean deprivation) were included to capture aggregate effects that may have a different impact to equivalent lower level effects (e.g., Firth et al., 2019; Saxon & Barkham, 2012).

This study was interested in determining the independent contribution of variables as much as possible, even if these partially mediated other effects. An exception to this was that we did not adjust for treatment duration, as we believed that potential bidirectional relationships with clinical effectiveness might confound interpretation.

Analysis

Multilevel models were used to analyze data. Multilevel models explicitly account for hierarchical clustering of variance, and can model random effects in addition to fixed effects (Hox, 2010; Snijders & Bosker, 2012). Cross-classified multilevel models were used to account for non-hierarchical clustering in the current dataset. For example, patients who live in the same neighborhood may receive treatment from different therapists, but equally one therapist may work across multiple neighborhoods, meaning therapists do not nest within neighborhoods nor vice versa. Similarly, one clinic covers multiple neighborhoods, but patients in some neighborhoods have the choice of multiple clinics. Cluster effects were calculated as the percentage of overall unexplained variance at a particular modeling level (e.g., the neighborhood effect in a model was the unexplained neighborhood level variance divided by total unexplained variance in that model, etc.) (Wampold & Owen, 2021). Our use of the word "effect" does not imply a causal interpretation – this variance could be produced in part by selection into neighborhoods and clinics as a result of unmeasured variables.

Data were analyzed using MLwiN 3.01 and 3.05 (Charlton et al., 2020). Multilevel models included up to 4 levels: patient, therapist, neighborhood, and clinic, and were estimated using Markov chain Monte Carlo (MCMC) estimation (Browne, 2017; Browne & Draper, 2006; Goldstein, 1989). Priors were generated using restricted iterative generalized least squares (RIGLS) estimation (continuous outcome models), or marginal quasi-likelihood (MQL) followed by predictive/penalized quasi-likelihood (PQL) estimation (logistic outcome models). MCMC chain length was set to exceed Nhats for tested parameters, to help ensure accuracy of coefficient and quantile estimates (Browne, 2017; Draper, 2008). All continuous variables were centered around their grand mean, and reference categories for categorical variables were set to the most common category. The therapist level in LI models was weighted according to the percentage of intervention that the primary PWP delivered (0-100%). The average primary PWP weighting was 85.8% (SD = 18.5%).

Models were constructed incrementally in two stages. First, variance components were tested (therapist, neighborhood, and clinic levels) without independent variables, to determine unadjusted clustering effects. Second, fixed main effects for independent variables were added and tested to see a) if they had significant associations with outcomes, and b) how much of the variance at each level was explained by those variables. Adjusted clustering effects are also presented.

Adjustment changes the interpretation of treatment outcome – it assumes that the variables adjusted for are not part of the effect of interest. However, in some cases these variables may be part of the effect of interest - this can be particularly relevant for cluster effects (even if those effects are in part selection effects, for example). As such, we have presented both the unadjusted and adjusted effects, which should be interpreted appropriately in context.

The study originally planned to test random slopes and corresponding cross-level interactions ([ANONYMISED FOR PEER REVIEW]). However, this limits and complicates interpretation of cluster effects in a way that was considered counter-productive for readers. As such, these stages of model development were not included in the final study analyses. Sensitivity analyses using these model specifications were conducted for rigor and completeness, and all results followed identical patterns (available on request from authors).

Backward elimination was used at each stage. For a coefficient to be retained required that its removal would result in an increase in the model's deviance information criterion (DIC; a measure of multilevel model fit that accounts for model complexity, Spiegelhalter et al., 2002), as well as significance of either a) Wald tests for random effects (acknowledging limitations; Browne, 2017) or b) MCMC credible intervals for fixed effects. Critical values for significance testing were adjusted using the Holm-Bonferroni method (Holm, 1979) at each stage, to correct for within-stage family-wise error rate due to multiple comparisons. Cluster effects were compared by calculating a full chain and associated credible intervals for the difference between each value in the originating MCMC chains.

Model assumptions were met, with the exception that the neighborhood and therapist residuals showed "tails" of smaller residuals at the extreme ends of outcome prediction. There was minimal skew apparent. Cohen's (1988) *d* effect sizes were also calculated.

When investigating variance explained by a variable or set of variables, the order in which variables are controlled for can radically affect how results are interpreted. In particular, variables adjusted for sequentially earlier can appear to explain relatively more variance than those adjusted for sequentially later. For this reason, we took two actions. First, we analysed the variance explained by deprivation factors both *before* and *after* controlling for other variables using traditional sequential approaches. Second, we developed visualizations to show the variance at each level that is *independently* explained by each variable or variable set, simultaneously controlling for all others. This avoids the problem of sequential privilege.

Results

Descriptive demographic and clinical statistics are shown in Table 1 and Table 2, respectively. A comparison with national data is included in Supplementary Material, indicating good representativeness of the sample overall. Both LI and HI samples had a mean age of 40, with 65-66% female patients. There were 79-82% patients from White backgrounds, 9-12% patients from ethnic minority backgrounds, and 9% patients in each sample with no recorded ethnicity. There were 55-60% employed patients, and 10-13% had a recorded disability (Table 1). In the HI sample patients attended a mean of 6.9 (*SD* = 4.8) sessions, with 89% of sessions delivered face-to-face. Pre-post Cohen's *d* treatment effect sizes were 0.69 (depression) and 0.72 (anxiety). In the LI sample patients attended a mean of 4.3 (*SD* = 2.5) sessions, with 49% delivered face-to-face. Average pre-post Cohen's *d* treatment effect sizes were 0.55 (depression) and 0.61 (anxiety).

Unadjusted Clustering Effects

We use the word unadjusted effects here to refer to clustering effects before controlling for independent variables (variance components models: first stage of analysis). Unadjusted therapist, neighborhood, and clinic effects across primary and secondary outcomes for high and low intensity treatments are shown in Figure 2. The primary outcome was post-treatment depression scores. For high intensity interventions, the model estimated an unadjusted therapist effect of 5.6% (Credible Interval 5.3% to 5.9%), an unadjusted neighborhood effect of 1.9% (CrI 1.8% to 2.0%), and an unadjusted clinic effect of 1.9% (CrI 1.3% to 2.8%). For low intensity interventions, the model estimated an unadjusted therapist effect of 5.4% (CrI = 5.2% to 5.7%), an unadjusted neighborhood effect of 2.2% (CrI = 2.1% to 2.3%), and an unadjusted clinic effect of 3.3% (CrI = 2.2% to 4.7%; Figure 2). As can be seen in Figure 2, unadjusted cluster effects for the recovery outcome were similar (6-7% for therapist effects, 1-2% for neighborhood effects, 3-5% for clinic effects). Post-hoc analyses using RCSI on a) the PHQ-9 and b) the GAD-7 as the outcome also found unadjusted cluster effects all within these same ranges (see Supplementary Material for comparison table).

In all models, best fit (lowest DIC and all Wald tests $p<.001$) was shown when four levels were modeled (patient, therapist, neighborhood, and clinic levels; full model specifications in Supplementary Material). This supports our first hypothesis.

Unadjusted cluster effects were compared between outcome models. A figure showing all comparisons is included in Supplementary Material. Clinic effects and neighborhood effects were typically significantly larger for low intensity treatment compared with high intensity treatment, whilst therapist effects were typically similar across both treatment intensities.

In the HI sample, residuals indicated an average 22% difference in predicted probability of recovery between therapists in the top quintile of predicted outcomes compared with therapists in the bottom quintile (26% difference for the LI sample). There was a 6% difference in predicted probability of recovery between top and bottom quintile neighborhoods (7% for the LI sample), and for there was an 18% difference between top and bottom quintile clinics (25% for the LI sample).

Adjusted Clustering Effects

Clustering effects were adjusted by controlling for all independent variables (both deprivation variables and control variables). In the high intensity treatment sample, variables significantly associated with lower post-treatment depression severity included lower baseline depression, anxiety, and functional impairment scores, being older, female, employed, White or Black (relative to other ethnic categories), not disabled, later treatment start date, not receiving prior low intensity treatment within the current referral, higher percentage of face-to-face sessions, less neighborhood income deprivation, less neighborhood crime deprivation, and higher neighborhood White ethnic density (all $p<.01$ and significant after Holm-Bonferroni correction; Table 3). Non-significant variables were neighborhood deprivation in the following domains: living, housing, education, employment, and health, as well as clinic mean deprivation and clinic white ethnic density. In the low intensity sample, the post-treatment depression model was highly similar - exceptions were that neighborhood education, living, and housing deprivation were also significant variables (all $p<.02$), and that neighborhood ethnic density and start date were not significant.

After controlling for all significant variables, the adjusted high intensity therapist effect was 4.06% (CrI 3.86% to 4.26%), neighborhood effect was 0.03% (0.003% to 0.08%), and clinic effect was 1.09% (0.72% to 1.61%). The adjusted low intensity therapist effect was 3.9% (CrI = 3.7% to 4.1%) neighborhood effect was 0.1% (CrI = 0.06% to 0.14%), and clinic effect was 2.0% (CrI = 3.7% to 4.1%; Figure 4).

Secondary outcome model specifications for each treatment type were generally highly comparable to their equivalent primary outcome model. The only differences between high intensity models were that gender was not significant in the recovery model, and neighborhood ethnic density was not significant in the anxiety or recovery models. The only differences between low intensity models were that neighborhood employment deprivation was significant in the anxiety model, and neighborhood education deprivation was not significant in the recovery model, but start date was. Adjusted cluster effects for primary and secondary outcome models are shown in Figure 3 (a table is also included in Supplementary Material). Post-hoc analyses of RCSI outcomes for the PHQ-9 and GAD-7 found almost identical adjusted cluster effects to those estimated for recovery outcomes (see Supplementary Material).

Explaining Cluster Effects

For the primary outcome (post-treatment depression severity) in the high intensity treatment sample, deprivation variables reduced the neighborhood effect from 1.8% to 0.2% if other variables were not controlled for (-1.67% change, CrI = -1.79% to 1.56%), or from 0.3% to 0.0% after controlling for other variables (-0.29% change, CrI = -0.36% to -0.22%). Similarly, in the low intensity treatment sample, deprivation variables reduced the neighborhood effect from 2.2% to 0.2% if other variables were not controlled for (-1.93% change, CrI = -2.04% to -1.82%), or from

0.5% to 0.1% after controlling for other variables (-0.41% change, CrI = -0.47% to -0.35%). Reduction in neighborhood variance (from which the neighborhood effect is derived) was similarly significant in all cases. In contrast, inclusion of deprivation variables did not significantly affect either the clinic variance or the size of clinic effect in all models, whether other variables were already controlled for or not (all p>.05). These results support our second hypothesis. A full table of results for primary and secondary outcomes is shown in Supplementary Material.

The *independent* ability of each variable or of set of variables to explain clustering at each level (by reducing unexplained variance at that level in the model, controlling for all other variables) is shown in Figure 4 for the high intensity depression outcome model. Figure 4 shows that while almost all of the neighborhood clustering variance could be explained by the included variables, approximately a fifth of the clinic effect could not be accounted for in this model. However, this figure shows that the majority of variance at most levels of clustering was not explainable by the independent contribution of any *single* variable, but was instead complex; i.e., variance was explained only by the shared effect of a number of variables, over and above their individual effects (as shown in Figure 4).

In the high intensity depression model, the majority of the neighborhood effect (55% of neighborhood variance) was only independently explained by the shared effect of deprivation and baseline severity variables, plus a further 11% independently explained by the shared effect of deprivation, baseline, and ethnicity variables, and a further 10% independently explained by the shared effect of deprivation variables alone. At the clinic level, the greatest percentage of clinic effect was also independently explained by the shared effect of deprivation and baseline severity variables (42%; Figure 4).

The low intensity depression model produced similar results (Figure 5), with the majority of neighborhood effect explained by the shared effect of deprivation and baseline severity (58%), and a further 11% independently explained by deprivation variables alone. The shared effect of deprivation and baseline severity variables only independently explained 18% of clinic variance.

Anxiety outcome models were very similar, with the shared effect of deprivation and baseline severity variables explaining the majority of neighborhood variance (between 53% and 58%), but only 17% of clinic variance (figures provided in Supplementary Material).

Discussion

This study is the first to our knowledge to statistically estimate neighborhood clustering effects on psychological intervention outcomes, and the first to simultaneously model neighborhood and clinic effects in this context. Neighborhood and clinic effects were found across both high and low intensity treatment types, for all outcomes (depression severity, anxiety severity, and clinical recovery). Neighborhood effects were comparable to or smaller than clinic effects, which in turn were smaller than therapist effects. Cluster effects were generally highly consistent across multiple outcomes.

Results suggest that patients living in some neighborhoods had systematically poorer treatment outcomes than others, but that this neighborhood effect could be completely explained by variables included in this study. Deprivation independently accounted for a significant proportion of neighborhood clustering, but the majority of neighborhood clustering could only be independently explained by a combined effect of baseline mental health severity and deprivation variables. Both groups of variables have been associated with clinical outcomes with relative consistency (Constantino et al., 2021). One explanation for the observed overlap of explained variance in this study could be a mediating effect. Theory and prior evidence suggest that deprivation increases risk of mental illness (Heinz et al., 2013; Johnson et al., 1999; Paul & Moser, 2009; Perry, 1996). A recent study further identified a specific psychological mindset that mediates the relationship between deprivation/SES and depression, characterized by low life satisfaction, low happiness, low optimism, and low resilience (Velten et al., 2021). It may therefore be important to consider deprivation as a causal factor, rather than simply a correlate or consequence of mental ill health.

In contrast, clinic effects were not significantly explained by deprivation variables. This was predicted, due to evidence of other organizational factors that may have a stronger clinic-level impact (Falkenström et al., 2018). However, results may also be influenced by statistical limitations. Broader credible intervals, due in part to the relatively smaller number of clinics (55 clinics versus 19,000 neighborhoods), may have reduced power at the clinic level. Further research into associations between deprivation and clinic level differences may therefore still be warranted. Ideally, such studies would also analyze more proximal clinic level variables, compared with the more distal measures possible for this study.

Larger cluster effects (at both neighborhood and clinic level) were detected for low intensity (compared with high intensity) treatment outcomes, even after controlling for predictor variables. This is clearest for recovery outcomes. A larger clinic effect for low intensity interventions suggests that organizational factors may have a greater relative influence on their successful delivery. Examples of factors that have been found to influence treatment outcomes include appropriate triage and intervention selection processes, length of wait before treatment, and organizational climate and culture (Clark et al., 2018; Falkenström et al., 2018; Gyani et al., 2013). Similarly, a larger neighborhood effect might suggest that low intensity interventions may rely more on a patient's local/neighborhood support network and resources to be successful.

A striking and unexpected (but specific and systematic) finding was that controlling for certain variables actually revealed a larger clinic effect than was apparent in the unadjusted effect – although this increased variability was itself explained by other variables in the final model. This finding suggests that it may be misleading to interpret unadjusted or adjusted clustering effects in isolation without also modelling and understanding the process by which relevant variables explain variance. For example, increased outcome variability between clinics caused by certain case-mix variables may be balanced out by different organizational approaches in response to these variables (e.g. additional training/support). This situation might yield a small unadjusted and/or adjusted clinic effect (due to the responsive efforts of the clinic), but could mask important contextual or operational differences that would be useful to understand and/or respond to. The approach used in this study to map out the variance explained by different variables could help counter this risk in future research. Non-linear relationships or ceiling effects etc. may also be useful to explore.

Based on the current study findings, particularly identifying neighborhood effects and the contributions of deprivation to clinical outcome, we encourage policymakers and commissioners to consider adjusting outcome targets for socio-clinical context. In the field of education, there is general consensus that so-called *value added* targets are preferable versus universal attainment targets, due in part to concerns that universal threshold targets (such as those affecting clinics in this study) may produce unintended incentives or consequences for outcomes (Prior et al., 2021; Ray et al., 2009). There are similar longstanding arguments in healthcare (Goldstein & Spiegelhalter, 1996). Despite this, the optimal specification of such contextual/value added targets continues to be debated (American Statistical Association, 2014; Prior et al., 2021; Ray et al., 2009).

It should of course be noted that the magnitude of cluster effects may change according to outcome used. For example, clinics achieving similar recovery rates (an example of attainment outcomes) may exhibit a different pattern of variability in pre-post change outcomes (an example of value added outcomes), etc.

Strengths and Limitations

The large sample enabled reliable examination of clustering effects at various levels, using a sophisticated multilevel modeling approach. The sample was representative of the population accessing common mental disorder treatment in England. As with many similar studies, these results indicate associations between variables but cannot directly infer causation. This extends to clustering effects themselves (a central premise of this study). For example, a proportion of clinic effects may be selection effects, rather than being related to how the clinic is managed (and similarly for therapist and neighborhood effects). The use of a large *N* routine dataset made this study feasible, but limited the variables available for inclusion in models (particularly at clinic and therapist levels). Unmeasured correlates such as expectancy, chronicity, or history of treatment may also demonstrate associations or causal links, or provide alternative interpretations to the current findings. Future studies could seek to develop understanding by integrating current findings with other prognostic factors. Clinic-level funding would be particularly valuable to include if possible. This study focused on IAPT clinics in part because the IAPT initiative aims to conform to a relatively standardized treatment framework. This helps to reduce confounding due to extraneous factors, but may mean that the clinic effects in this study are relatively conservative compared with some treatment contexts. No post-treatment follow-up data was available, preventing investigation of longer term differences in outcomes.

Conclusions

This study found that psychological treatment effectiveness varies systematically according to the neighborhood where a patient lives, as well as the clinic where they are treated. A shared effect of deprivation and baseline symptom severity was key in explaining the neighborhood (but not clinic) effect. Clinic and neighborhood effects were larger for low intensity treatment compared with traditional high intensity treatment. The effect of socioeconomic deprivation on outcomes may also be broader for low intensity treatment (more domains of deprivation predicted outcomes).

Identifying variables capable of more fully explaining the clinic effect is a priority for future research (as has been the case for therapist effects research for a number of years). In addition, research should aim to focus on understanding the causal mechanisms that might further explain and clarify clustering effects at both neighborhood and clinic levels.

Data Transparency

There are no other studies from the same dataset. A summary of findings from this study using similar model specifications was presented at the Society for Psychotherapy Research 52nd Annual International Meeting (2021, June 23-26).

References

Allison, P. (2001). *Missing Data (Quantitative Applications in the Social Sciences)*. Sage.

American Psychological Association Task Force on Socioeconomic Status. (2007). *Report of the APA Task Force on Socioeconomic Status*. American Psychological Association. https://www.apa.org/pi/ses/resources/publications/task-force-2006.pdf

American Statistical Association. (2014). *ASA Statement on Using Value-Added Models for Educational Assessment*. https://www.amstat.org//asa/files/pdfs/POL-ASAVAM-Statement.pdf

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258-297). Wiley.

Brand, C. A., Barker, A. L., Morello, R. T., Vitale, M. R., Evans, S. M., Scott, I. A., Stoelwinder, J. U., & Cameron, P. A. (2012). A review of hospital characteristics associated with improved performance. *International Journal for Quality in Health Care*, *24*(5), 483-494. https://doi.org/10.1093/intqhc/mzs044

Browne, W. J. (2017). *MCMC Estimation in MLwiN*. Centre for Multilevel Modelling, University of Bristol. http://www.bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/3-00/mcmc-web.pdf

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*(3), 473-513. https://doi.org/10.1214/06-ba117

Cameron, I. M., Crawford, J. R., Lawton, K., & Reid, I. C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice*, *58*(546), 32-36. https://doi.org/10.3399/bjgp08X263794

Charlton, C., Rasbash, J., Browne, W. J., Healy, M., & Cameron, B. (2020). *MLwiN* (Version 3.05) Centre for Multilevel Modelling, University of Bristol.

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, *23*(4), 318-327. https://doi.org/10.3109/09540261.2011.606803

Clark, D. M. (2018). Realizing the Mass Public Benefit of Evidence-Based Psychological Therapies: The IAPT Program. *Annual Review of Clinical Psychology, Vol 14*, *14*, 159-183. https://doi.org/10.1146/annurev-clinpsy-050817084833

Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *Lancet*, *391*(10121), 679-686. https://doi.org/10.1016/s0140-6736(17)32133-5

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Earlbaum Associates.

Constantino, M. J., Boswell, J. F., & Coyne, A. C. (2021). Patient, therapist, and relational factors. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergen and Garfield's Handbook of Psychotherapy and Behavior Change* (7th ed.). Wiley.

Delgadillo, J., Farnfield, A., & North, A. (2018). Social inequalities in the demand, supply and utilisation of psychological treatment. *Counselling and Psychotherapy Research*, *18*, 114-121. https://doi.org/10.1002/capr.12169

Draper, D. (2008). Bayesian multilevel analysis and MCMC. In J. d. Leeuw & E. Meijer (Eds.), *Handbook of Multilevel Analysis*. Springer.

Falkenström, F., Grant, J., & Holmqvist, R. (2018). Review of organizational effects on the outcome of mental health treatments. *Psychotherapy Research*, *28*(1), 76-90. https://doi.org/10.1080/10503307.2016.1158883

Finegan, M., Firth, N., & Delgadillo, J. (2020). Adverse impact of neighbourhood socioeconomic deprivation on psychological treatment outcomes: the role of area-level income and crime. *Psychotherapy Research*, *30*(4), 546-554. https://doi.org/10.1080/10503307.2019.1649500

Finegan, M., Firth, N., Wojnarowski, C., & Delgadillo, J. (2018). Associations between socioeconomic status and psychological therapy outcomes: A systematic review and meta-analysis. *Depression and Anxiety*, *35*(6), 560-573. https://doi.org/10.1002/da.22765

Firth, N., Barkham, M., Kellett, S., & Saxon, D. (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: A multilevel modelling analysis. *Behaviour Research and Therapy*, *69*, 54-62. https://doi.org/10.1016/j.brat.2015.04.001

Firth, N., Saxon, D., Stiles, W. B., & Barkham, M. (2019). Therapist and clinic effects in psychotherapy: a three-level model of outcome variability. *Journal of Consulting and Clinical Psychology*, *87*, 345-356. https://doi.org/10.1037/ccp0000388

Freedman, D., & Woods, G. W. (2013). Neighborhood Effects, Mental Illness and Criminal Behavior: A Review. *Journal of politics and law*, *6*(3), 1-16. https://doi.org/https://doi.org/10.5539/jpl.v6n3p1

Goldstein, H. (1989). Restricted unbiased iterative generalized least-squares estimation. *Biometrika*, *76*(3), 622-623. https://doi.org/10.1093/biomet/76.3.622

Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. Journal of the Royal Statistical Society Series a-Statistics in Society, 159, 385-409. https://doi.org/10.2307/2983325

Green, H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT Psychological Wellbeing Practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy*, *63*, 43-54. https://doi.org/10.1016/j.brat.2014.08.009

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, *51*(9), 597-606. https://doi.org/10.1016/j.brat.2013.06.004

Heinz, A., Deserno, L., & Reininghaus, U. (2013). Urbanicity, social adversity and psychosis. *World Psychiatry*, *12*(3), 187-197. https://doi.org/10.1002/wps.20056

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65-70.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge.

Jacobson, N. S., & Truax, P. (1991). Clinical significance - A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12-19. https://doi.org/10.1037//0022-006x.59.1.12

Johns, R., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: An update and refinement to Baldwin and Imel's (2013) review. *Clinical Psychology Review*, *67*, 78–93. https://doi.org/10.1016/j.cpr.2018.08.004

Johnson, J. G., Cohen, P., Dohrenwend, B. P., Link, B. G., & Brook, J. S. (1999). A longitudinal investigation of social causation and social selection processes involved in the association between socioeconomic status and psychiatric disorders. *Journal of Abnormal Psychology*, *108*(3), 490-499. https://doi.org/10.1037/0021-843x.108.3.490

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9 - Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Longevity Science Panel. (2018). *Life expectancy: Is the socio-economic gap narrowing?* http://www.longevitypanel.co.uk/_files/LSP_Report.pdf

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *General Hospital Psychiatry*, *28*(1), 71-77. https://doi.org/10.1016/j.genhosppsych.2005.07.003

Ministry of Housing, Communities & Local Government. (2019). *English Indices of Deprivation 2019*.

Moineddin, R., Matheson, F. I., & Glazier, R. H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, *7*, 10, Article 34. https://doi.org/10.1186/1471-2288-7-34

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *British Journal of Psychiatry*, *180*, 461-464. https://doi.org/10.1192/bjp.180.5.461

Muntaner, C., Eaton, W. W., Miech, R., & O'Campo, P. (2004). Socioeconomic position and major mental disorders. *Epidemiologic Reviews*, *26*, 53-62. https://doi.org/10.1093/epirev/mxh001

NHS Digital. (2021). *Lower Layer Super Output Area*. https://datadictionary.nhs.uk/nhs_business_definitions/lower_layer_super_output_area.html

NHS England, & NHS Improvement. (2017). *Developing an outcomes-based payment approach for IAPT services*. https://improvement.nhs.uk/documents/661/IAPT_Payment_Guidance.pdf

O'Donoghue, B., Roche, E., & Lane, A. (2016). Neighbourhood level social deprivation and the risk of psychotic disorders: a systematic review. *Social Psychiatry and Psychiatric Epidemiology*, *51*(7), 941-950. https://doi.org/10.1007/s00127-016-1233-4

Office for National Statistics. (2016). *Ethnic group, national identity and religion*. https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion#ethnic-group

Office for National Statistics. (n.d.). *2011 Census*. https://www.ons.gov.uk/census/2011census

Paul, K. I., & Moser, K. (2009). Unemployment impairs mental health: Meta-analyses. *Journal of Vocational Behavior*, *74*(3), 264-282. https://doi.org/10.1016/j.jvb.2009.01.001

Pepinsky, T. B. (2018). A Note on Listwise Deletion versus Multiple Imputation. *Political Analysis*, *26*(4), 480-488. https://doi.org/https://doi.org/10.1017/pan.2018.18

Perry, M. J. (1996). The relationship between social class and mental disorder. *The Journal of Primary Prevention*, *17*(1), 17-30. https://doi.org/10.1007/bf02262736

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health*, *55*(2), 111-122. https://doi.org/10.1136/jech.55.2.111

Poole, R., Higgo, R., & Robinson, C. A. (2014). *Mental Health and Poverty*. Cambridge University Press.

Poverty and Social Exclusion UK. (2016, January 21). *Deprivation and poverty*. http://www.poverty.ac.uk/definitions-poverty/deprivation-and-poverty

Prior, L., Jerrim, J., Thomson, D., & Leckie, G. (2021). A review and evaluation of secondary school accountability in England: Statistical strengths, weaknesses and challenges for 'Progress 8' raised by COVID-19. *Review of Education,* 9(3). e3299. https://doi.org/10.1002/rev3.3299

Ray, A., McCormack, T., & Evans, H. (2009). Value added in English schools. *Education Finance and Policy*, 4(4), 415-438. https://doi.org/10.1162/edfp.2009.4.4.415

Richardson, R., Westley, T., Gariepy, G., Austin, N., & Nandi, A. (2015). Neighborhood socioeconomic conditions and depression: a systematic review and meta-analysis. *Social Psychiatry and Psychiatric Epidemiology*, *50*(11), 1641-1656. https://doi.org/10.1007/s00127-015-1092-4

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, *80*, 535-546. https://doi.org/10.1037/a0028898

Schiefele, A.-K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., Kopta, M., Schulte, D., Saxon, D., Nielsen, S. L., & Lambert, M. J. (2017). Reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*, 598–613. https://doi.org/10.1007/s10488-016-0736-3

Silva, M., Loureiro, A., & Cardoso, G. (2016). Social determinants of mental health: a review of the evidence. *European Journal of Psychiatry*, *30*(4), 259-292.

Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling. In (2nd ed.). London: Sage Publishers.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *64*, 583-616. https://doi.org/10.1111/1467-9868.00353

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder - The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

The National Collaborating Centre for Mental Health. (2021). *The Improving Access to Psychological Therapies Manual*. https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual-v5.pdf

Townsend, P. (1979). *Poverty in the United Kingdom*. Allen Lane and Penguin Books.

Townsend, P. (1987). Deprivation. *Journal of Social Policy*, *16*, 125-146. https://doi.org/10.1017/s0047279400020341

Velten, J., Scholten, S., Brailovskaia, J., & Margraf, J. (2021). Psychometric properties of the S-Scale: Assessing a psychological mindset that mediates the relationship between socioeconomic status and depression. *PloS one*, *16*(10), e0258333. https://doi.org/10.1371/journal.pone.0258333

Wampold, B. E., & Owen, J. (2021). Therapist effects: History, methods, magnitude, and characteristics of effective therapists. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergen and Garfield's Handbook of Psychotherapy and Behavior Change* (7th ed.). Wiley.

Webb, C. J. R., & Bywaters, P. (2018). Austerity, rationing and inequity: trends in children's and young peoples' services expenditure in England between 2010 and 2015. *Local Government Studies*, *44*(3), 391-415. https://doi.org/https://doi.org/10.1080/03003930.2018.1430028

Wilson, W. J. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

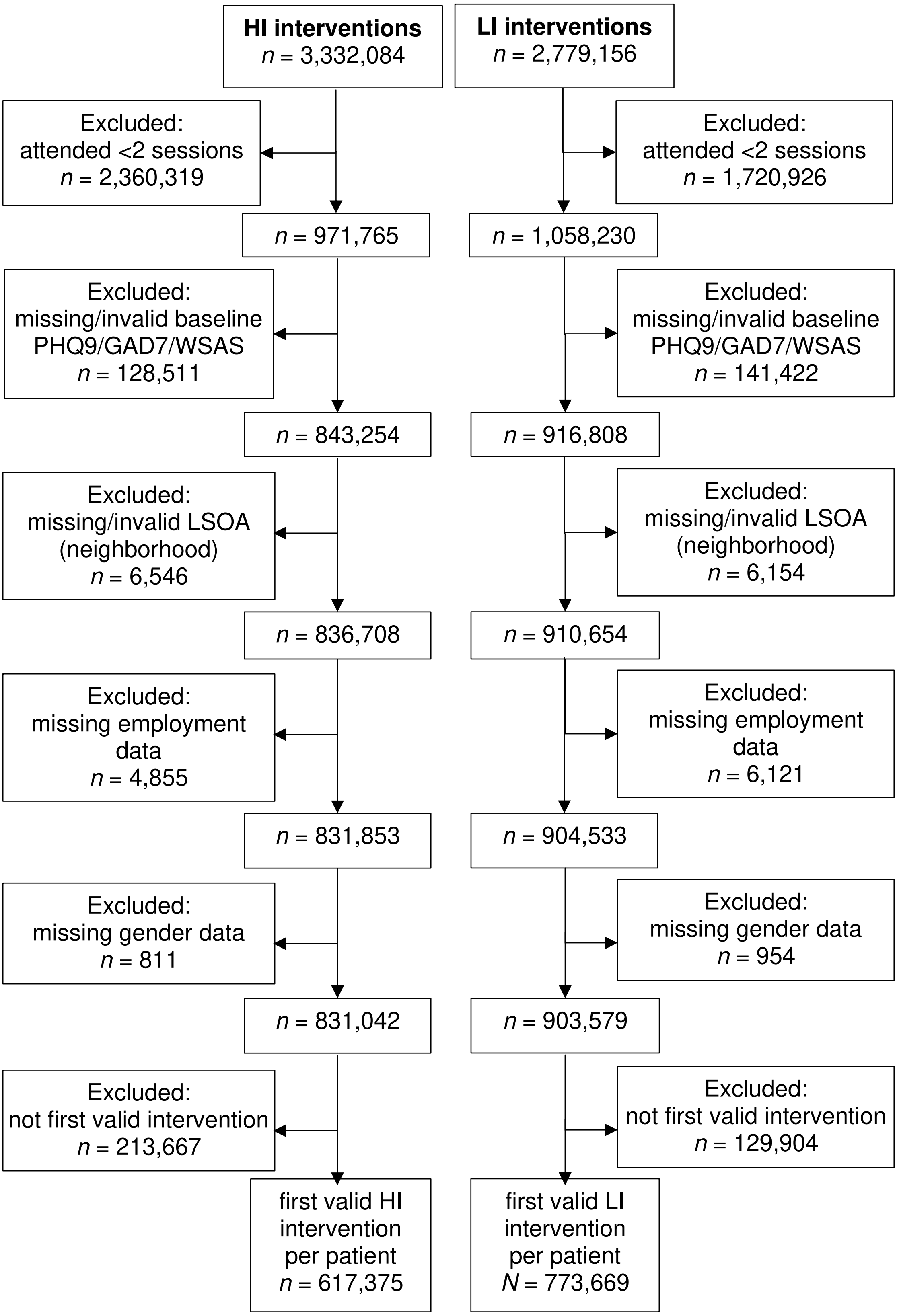

Figure 1. Sample inclusion flowchart

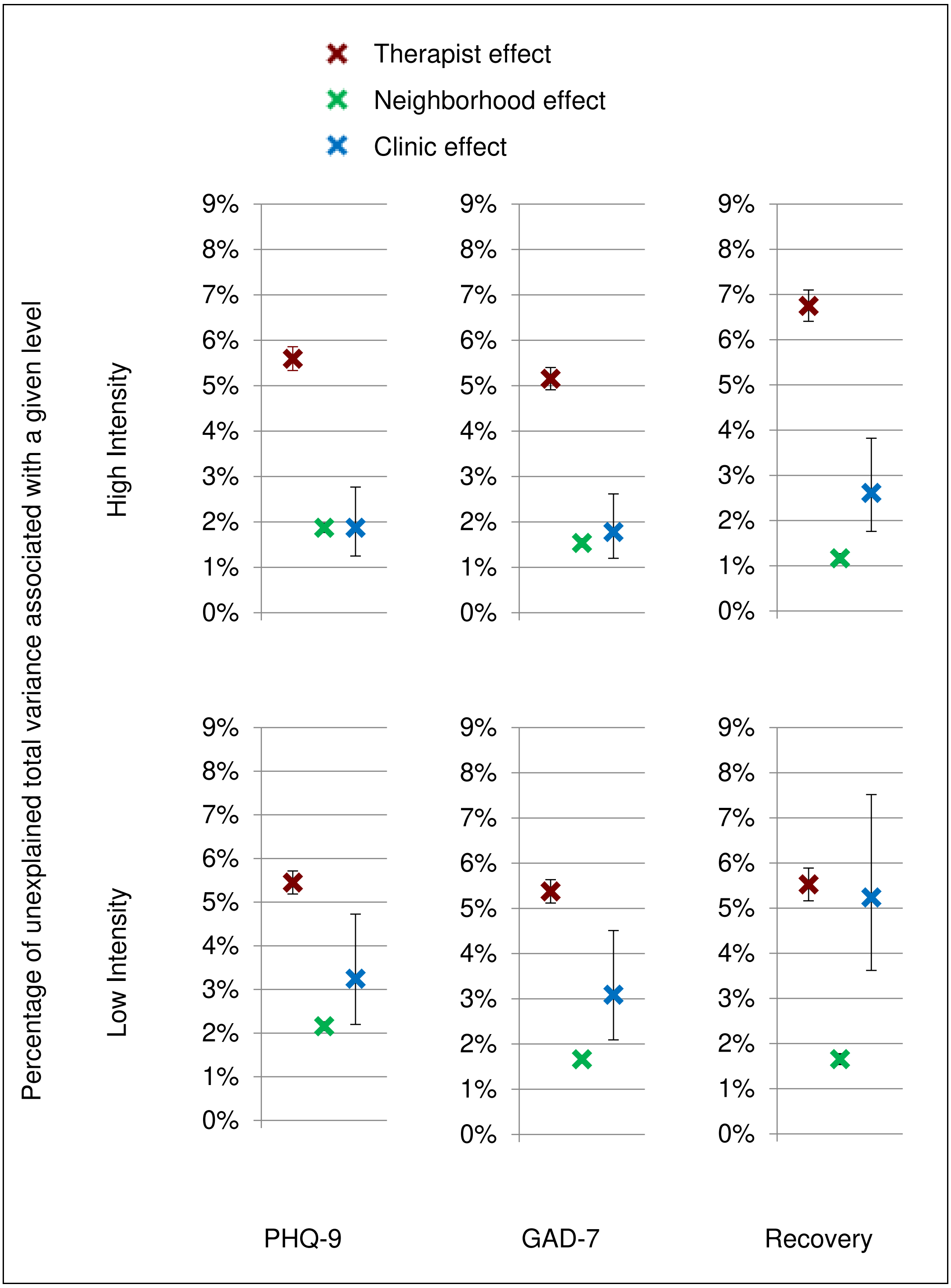

Figure 2. Unadjusted cluster effects across different treatments and outcome variables. Clustering effects are calculated as the percentage of total unexplained outcome variance in a model that is associated with the relevant level of clustering.

GAD-7 = Generalized Anxiety Disorder-7, PHQ-9 = Patient Health Questionnaire

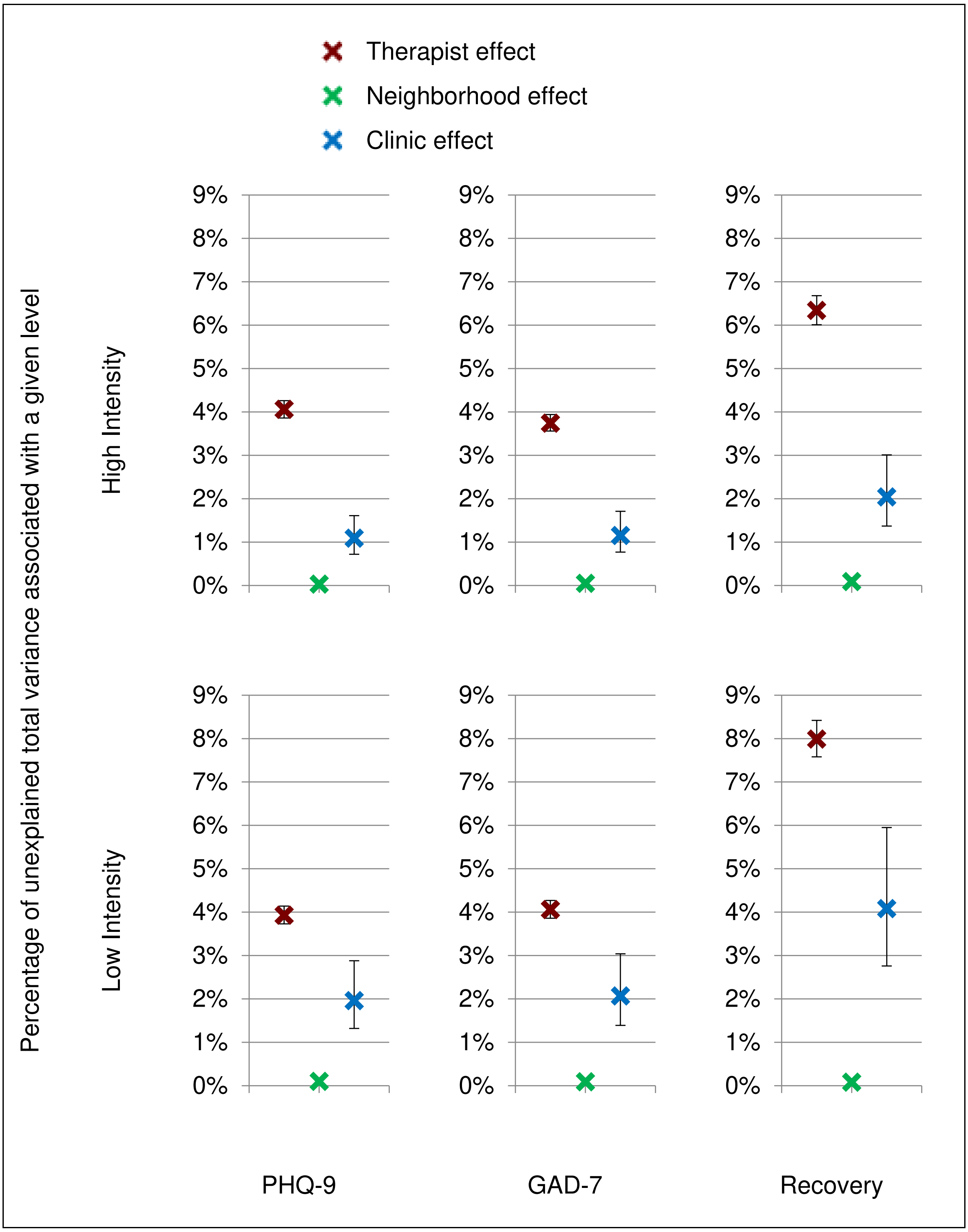

Figure 3. Adjusted cluster effects across different treatments and outcome variables after accounting for explanatory and control variables. Clustering effects are calculated as the percentage of total unexplained outcome variance in a model that is associated with the relevant level of clustering.

GAD-7 = Generalized Anxiety Disorder-7, PHQ-9 = Patient Health Questionnaire

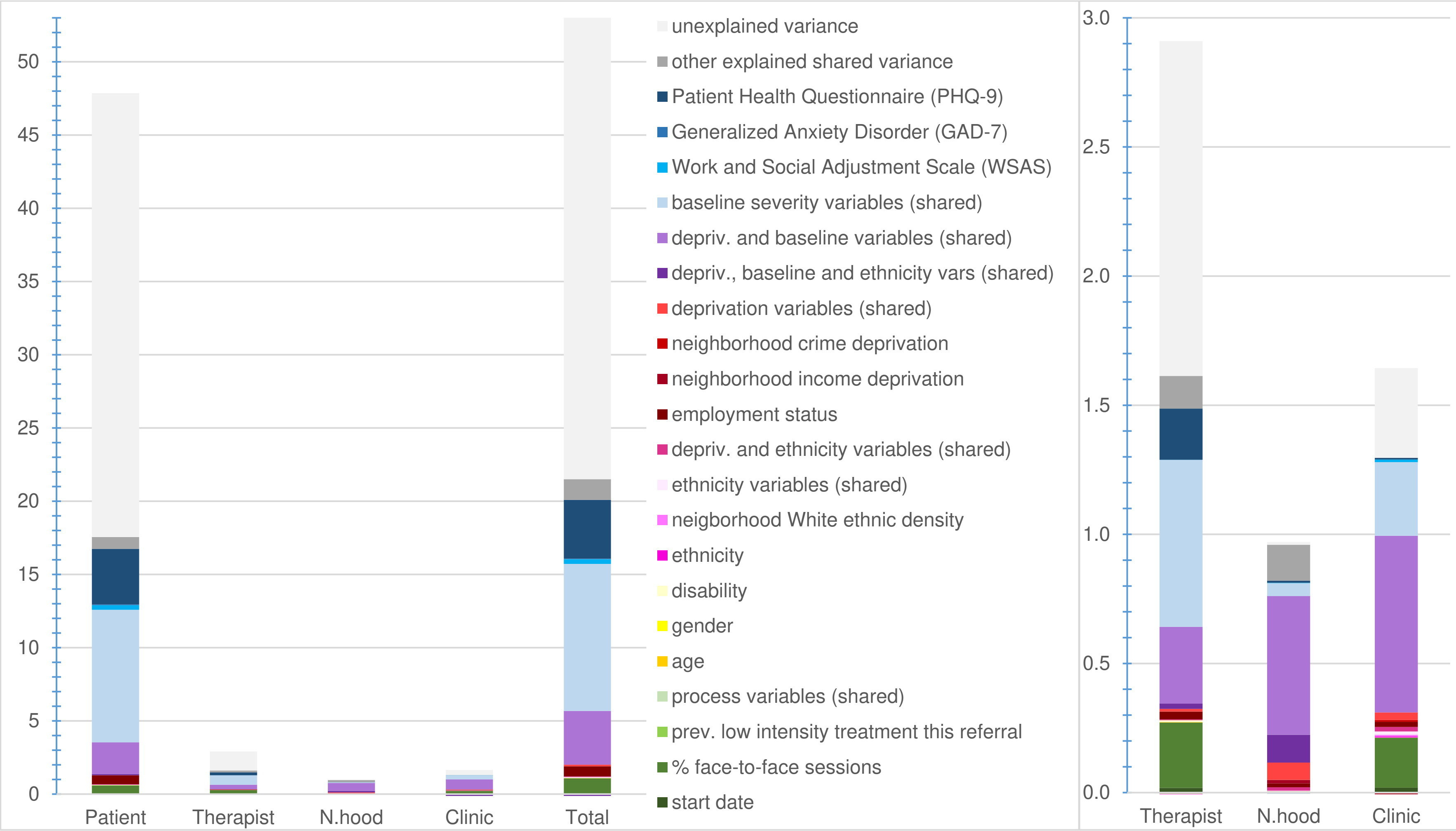

Figure 4. High intensity depression outcome variance (clustering) at each level, showing the variance independently explained by individual variables, explained residually by groups of variables, and unexplained by any variables. The right hand pane shows a close-up of the therapist, neighborhood, and clinic levels. Each type of variable is shown in a different color (e.g. baseline severity = blues, deprivation = reds, ethnicity = pinks, other demographic = yellows, process variables = greens). Purples indicate more complex shared effects.

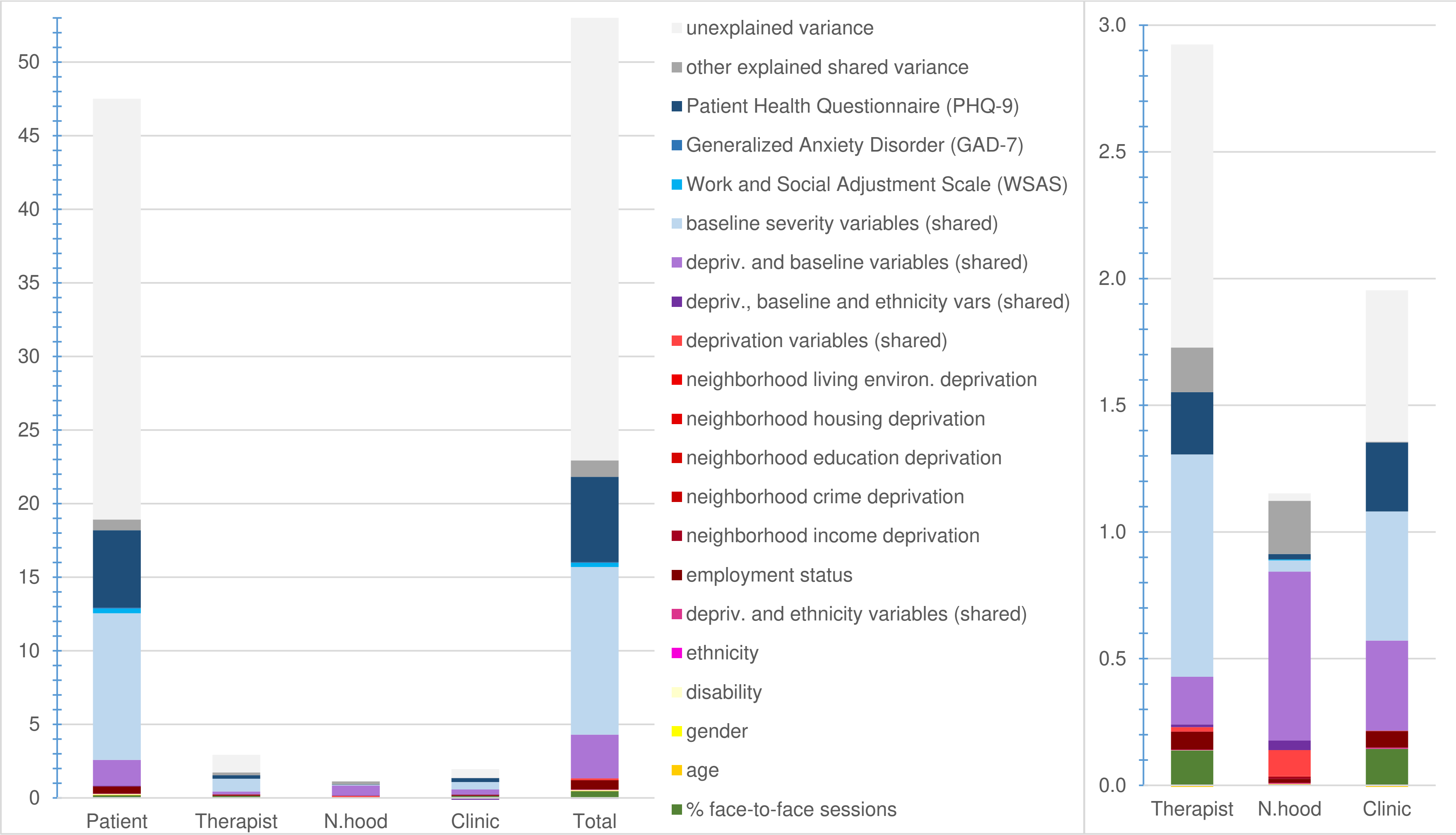

Figure 5. Low intensity depression outcome variance (clustering) at each level, showing the variance independently explained by individual variables, explained residually by groups of variables, and unexplained by any variables. The right hand pane shows a close-up of the therapist, neighborhood, and clinic levels. Each type of variable is shown in a different color (e.g. baseline severity = blues, deprivation = reds, ethnicity = pinks, other demographic = yellows, process variables = greens). Purples indicate more complex shared effects.

Table 1. *Descriptive demographic statistics for high and low intensity samples*

	Mean (SD), or percentage of sample			
	High Intensity sample		Low Intensity sample	
Mean Age	39.7	(14.2)	39.9	(15.0)
Female	66.1%		64.6%	
Ethnicity:				
White	79.0%		82.4%	
Mixed	2.1%		1.6%	
Asian	5.0%		4.1%	
Black	3.2%		2.2%	
Other	1.6%		1.1%	
not-recorded	9.1%		8.5%	
Employment status:				
employed	54.6%		60.2%	
retired	7.1%		8.2%	
volunteer	0.2%		0.2%	
student	4.4%		4.9%	
homemaker	7.0%		5.9%	
unemployed-no-benefits	0.7%		0.7%	
unemployed-seeking	17.8%		14.1%	
unemployed-benefits	8.2%		5.7%	
Disability status:				
no disability	64.8%		67.5%	
disability not recorded	22.5%		22.7%	
has disability	12.7%		9.9%	
Mean IMD-income p.c.	44.5	(26.6)	47.2	(26.9)
Mean IMD-employment p.c.	45.0	(27.3)	47.0	(27.6)
Mean IMD-education p.c.	48.5	(27.5)	48.4	(27.4)
Mean IMD-health p.c.	46.3	(27.4)	47.6	(27.7)
Mean IMD-crime p.c.	51.4	(28.7)	53.9	(28.9)
Mean IMD-housing p.c.	46.7	(29.9)	48.7	(29.4)
Mean IMD-living p.c.	46.2	(30.4)	49.1	(30.9)
Mean clinic mean IMD p.c.	39.7	(17.3)	43.0	(17.8)
Mean n.hood W-ED	84.5%	(21.1%)	87.8%	(18.3%)
Mean clinic W-ED	84.3%	(19.7%)	88.0%	(16.8%)

IMD = index of multiple deprivation. p.c. = percentile. W-ED = White ethnic density.

Table 2. *Descriptive clinical statistics for high and low intensity samples*

	Mean (SD), or percentage of sample	
	High Intensity sample	Low Intensity sample
Mean first appointment date	Mar-2015 (992 days)	Jun-2015 (1022 days)
Mean discharge date	Aug-2015 (986 days)	Oct-2015 (1021 days)
Previous LI this referral [a]	20.3%	-
Mean sessions scheduled	8.8 (5.8)	5.5 (3.2)
Mean sessions attended	6.9 (4.8)	4.3 (2.5)
Mean latest session with recorded outcome scores	6.8 (4.8)	4.1 (2.6)
Mean face-to-face contact	88.6% (26.7%)	49.4% (40.4%)
Mean pre-intervention PHQ-9	14.7 (6.5)	14.3 (6.3)
Mean post-intervention PHQ-9	10.0 (7.2)	10.6 (7.2)
Mean PHQ-9 change	-4.7 (6.3)	-3.7 (6.0)
Mean pre-intervention GAD-7	13.1 (5.4)	13.0 (5.3)
Mean post-intervention GAD-7	8.9 (6.2)	9.5 (6.2)
Mean GAD-7 change	-4.2 (5.7)	-3.5 (5.5)
Mean pre-intervention WSAS	19.0 (9.8)	18.3 (9.6)
Mean post-intervention WSAS	14.1 (10.5)	14.3 (10.4)
Mean WSAS change	-4.9 (8.8)	-4.0 (8.6)

[a] 2+ sessions

GAD-7 = Generalized Anxiety Disorder-7, LI = low intensity treatment, PHQ-9 = Patient Health Questionnaire, WSAS = Work and Social Adjustment Scale

Table 3. *Final 4-classification model specification for each outcome model.*

Variable	β estimate (and significance)					
	HI depression	LI depression	HI anxiety	LI anxiety	HI recovery†	LI recovery†
Baseline PHQ-9	0.493***	0.558***	0.138***	0.133***	-0.075***	-0.076***
Baseline GAD-7	0.059***	0.060***	0.399***	0.474***	-0.063***	-0.062***
Baseline WSAS	0.077***	0.078***	0.053***	0.052***	-0.022***	-0.023***
Age	-0.006***	-0.016***	-0.009***	-0.020***	0.007***	0.012***
Gender						
female (reference)						
male	-0.067***	-0.064***	-0.042**	-0.152***	-	0.044***
Employment:						
employed (reference)						
retired	0.303***	0.322***	0.304***	0.322***	-0.202***	-0.219***
other-role	0.826***	0.721***	0.709***	0.608***	-0.286***	-0.261***
unemp-nobenefits	1.647***	1.459***	1.360***	1.084***	-0.522***	-0.491***
unemp-seeking	1.954***	1.813***	1.575***	1.444***	-0.623***	-0.602***
unemp-benefits	2.239***	2.357***	1.832***	1.837***	-0.710***	-0.792***
Ethnicity:						
White (reference)						
Mixed	0.212***	0.229***	0.167***	0.180***	-0.069**	-0.069**
Asian	0.470***	0.376***	0.494***	0.376***	-0.130***	-0.078***
Black	-0.105*	0.010	-0.080*	-0.002	0.030	-0.006
Other	0.664***	0.545***	0.654***	0.458***	-0.194***	-0.094***
not recorded	0.030	0.124***	0.011	0.078***	-0.020	-0.042***
Disability:						
no (reference)						
yes	0.642***	0.846***	0.438***	0.572***	-0.177***	-0.266***
not recorded	0.142***	-0.042*	0.083***	-0.081***	-0.024*	0.089***
% Face to Face	-3.691***	-1.491***	-3.254***	-1.397***	1.772***	0.665***
Start Date	-1E-04***	-	-1E-04***	-	-1E-04***	7 E-05***
Previous LI	0.160***	n/a	0.124***	n/a	-0.048***	n/a

IMD-Income	-0.619***	-0.670***	-0.526***	-0.401***	0.181***	0.279***
IMD-Crime	-0.265***	-0.244***	-0.216***	-0.200***	0.106***	0.089***
IMD-Housing	-	-0.087**	-	-0.072**	-	0.031*
IMD-Education	-	-0.178***	-	-0.119**	-	-
IMD-Living	-	-0.069*	-	-0.064**	-	0.032**
IMD-Employment	-	-	-	-0.191*	-	-
IMD-Health	-	-	-	-	-	-
N.hood W-ED	-0.189***	-	-	-	-	-
Clinic mean IMD	-	-	-	-	-	-
Clinic W-ED	-	-	-	-	-	-

GAD-7 = Generalized Anxiety Disorder-7, HI = high intensity, IMD = neighborhood index of multiple deprivation, LI = low intensity, PHQ-9 = Patient Health Questionnaire, W-ED = White ethnic density, WSAS = Work and Social Adjustment Scale.

† recovery coefficient signs (+/-) are opposite to those for other outcomes, as recovery is inversely associated with post-treatment severity. Direction of effects are therefore clinically consistent.

* p < .05, ** p < .01, *** p < .001