Article:

Firth, N. orcid.org/0000-0003-1984-6869, Saxon, D. orcid.org/0000-0002-9753-8477, Stiles, W.B. et al. (1 more author) (2020) Therapist effects vary significantly across psychological treatment care sectors. Clinical Psychology & Psychotherapy. cpp.2461. ISSN 1063-3995

https://doi.org/10.1002/cpp.2461

Received: 10 December 2019 | Revised: 25 February 2020 | Accepted: 16 April 2020
DOI: 10.1002/cpp.2461

RESEARCH ARTICLE

WILEY

Therapist effects vary significantly across psychological treatment care sectors

Nick Firth[1] | David Saxon[2] | William B. Stiles[3,4] | Michael Barkham[2]

[1]School of Health and Related Research, University of Sheffield, Sheffield, UK
[2]Clinical Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK
[3]Department of Psychology, Miami University, Oxford, OH, USA
[4]Department of Psychology, Appalachian State University, Boone, NC, USA

Correspondence
Nick Firth, School of Health and Related Research, University of Sheffield, Sheffield S1 4DA, UK.
Email: nick.firth@gmail.com

Abstract

Psychological intervention outcomes depend in part on the therapist who provides the intervention (a therapist effect). However, recent reviews suggest that therapist effects may vary as a function of the context in which care is provided and therefore should not be generalized beyond that context. This study statistically analysed therapist effect differences between care sectors delivering psychological interventions. The sample comprised routine clinical data from 26,814 patients (69% female; mean age 38) and 466 therapists in five care sectors: *primary care*, *secondary care*, *university*, *voluntary*, and *workplace*. Therapist effects were analysed using multilevel models and Markov chain Monte Carlo credible intervals. The therapist effect was significantly larger in primary care (8.4%) than in any other sector (1.1%–2.3%) except secondary care (4.1%), after controlling for explanatory baseline and process variables as well as accounting for differences between clinics. There were no other significant differences detected between care sectors. These findings support the hypothesis that differences in effectiveness between therapists vary depending on the context in which psychological treatment is provided. Differences in relative therapist impact can vary by a factor of 4–8 across treatment sectors. This should be considered in the application of research evidence, treatment planning, and the design and delivery of psychological care provision.

KEYWORDS

multilevel, primary care, psychological therapy, secondary care, therapist effect, third sector

1 | INTRODUCTION

In the psychological intervention literature, therapist effects refer to systematic variability among therapists regarding patient outcomes, independent of patient, and treatment characteristics (Barkham, Lutz, Lambert, & Saxon, 2017). Evidence suggests approximately 5% of variance in outcome is associated with the therapist (Baldwin & Imel, 2013; Johns, Barkham, Kellett, & Saxon, 2019). Both of these reviews of therapist effects indicated that study heterogeneity can impact on the measurement of therapist effects. Johns et al.'s (2019) recent review drew particular attention to methodological factors such as study design (randomized controlled trial versus practice-based study), complexity of outcome measure, use of a reliable and validated outcome measure, and sufficient sample size. In order to reduce methodological confounding and derive reliable effects estimates, they recommend that future therapist effects research uses homogenous, practice-based designs, and multilevel analysis techniques.

Regardless of methodology, therapist effects may themselves vary as a function of care context, such that therapist effects should be considered only with respect to their specific corresponding care context. Johns et al. (2019) noted this problem but were unable to perform statistical comparisons with the data available from the published literature.

In the United Kingdom (as in many other countries), psychological interventions are delivered within a number of different care contexts, known as sectors. Sectors differ in how they are designed to deliver care. For example, primary care clinics typically provide brief, front-line interventions to a wide range of individuals with mild-to-moderate common mental health conditions. In contrast, secondary and tertiary sectors provide increasingly specialized and intensive interventions, to patients who often experience high problem complexity and/or risk of suicide or self-harm. Some sectors serve populations that may be highly circumscribed or distinct regarding demographic characteristics. For example, the university counselling sector will, in the main, offer care to younger, well-educated adults, and patients accessing the workplace counselling sector will, by definition, almost all be currently employed, compared with other sectors. Therapists may have different professional backgrounds, and other organizational factors can vary, particularly between private and public organizations (e.g., workplace policies, salary, and career progression). These differences at the patient, therapist, treatment, and organizational level may affect the relative contribution of the therapist to clinical variability. For example, greater patient symptom severity has been found to be associated with a larger therapist effect (Saxon & Barkham, 2012). These contextual questions are compounded by a relative lack of research investigating care contexts other than mainstream care providers (such as treatment provided by voluntary organizations or universities/colleges).

The current study aimed to use a large data set to address this question within a single cohesive investigation of therapist effects differences across five UK care sectors: primary care, secondary care, voluntary, university counselling, and workplace counselling. In doing so, it also sought to address a number of confounding methodological factors identified by Johns et al. (2019) in previous research—for example, by using a single statistical approach, a single, reliable, and well-validated outcome measure, and consistent case-mix variables, as well as controlling for clinic effects (Firth, Saxon, Stiles, & Barkham, 2019). Investigating variability in therapist effects contributes to addressing the question of why we observe differences in effectiveness between therapists and may thereby suggest how outcomes may potentially be improved.

Key Practitioner Message

- Psychological intervention outcomes vary according to the therapist providing the intervention (a therapist effect).
- Therapist allocation appears to be more important for outcome in some types of care context compared with others. Therapist effects cannot be assumed to be universally applicable.
- Therapist effects were 4–8 times greater in primary care, compared with university, voluntary, and workplace settings.
- The context-specific nature of therapist effects and therefore the variable association between therapist and patient outcome should be considered in psychological care planning and delivery.

2 | METHOD

2.1 | Measures

The primary outcome measure was the patient's post-therapy score on the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Evans et al., 2002), a 34-item measure of psychological distress comprising wellbeing, symptom severity, functioning, and risk subscales. Risk in this case refers to the patient's day-to-day risk to self and risk to others. Scores range from 0 to 40, with higher scores indicating greater distress. Scores above 10 indicate clinical distress (Connell et al., 2007). A valid CORE-OM score requires scores on at least 31 of 34 items (90% of items completed). Psychometric properties demonstrated (Barkham et al., 2010) include internal consistency of α = 0.93–0.95 (Barkham, Gilbert, Connell, Marshall, & Twigg, 2005), test–retest reliability of .88 (with outpatients at 1-month intervals; Barkham, Mullin, Leach, Stiles, & Lucock, 2007), and strong convergent validity (Cahill et al., 2006; Connell et al., 2007). The CORE-OM was completed by patients as part of routine practice.

Control variables were also part of the standard CORE data set, collected as part of routine practice using the CORE Assessment form. Control variables included were as follows: pre-therapy CORE-OM nonrisk severity and risk severity scores, patient age and employment status (employed, other role, and not employed), number of sessions attended and percentage of offered sessions attended, therapy frequency (more than weekly, weekly, less than weekly, and no fixed frequency), and a variable for clinics.

Pre-intervention patient severity is one of the strongest predictors of patient outcome, as have been found to be associated with therapist effect size (Bohart & Greaves Wade, 2013; Hamilton & Dobson, 2002; Saxon & Barkham, 2012). Patients' CORE-OM scores can be split into risk (to self and others) scores and nonrisk (e.g., wellbeing, symptoms, and functioning) scores. This was applied in the current study, given previous research finding that CORE-OM nonrisk scores were associated with therapist effect size (Saxon & Barkham, 2012), whereas risk scores predicted patient outcome but were not associated with the size of therapist effect. Both of these scores were transformed to range from 0 to 40, for consistency.

Patient age and employment status, and attendance of sessions also typically predict patient outcome. The number of sessions attended, attendance rate, and therapy frequency was also included in order to control for expected variation between sectors in care provision. For example, patients in voluntary sector and secondary care clinics have been found to attend an average of 13–16 sessions, compared with primary care, university, and workplace sector who attended an average of 6–7 sessions. As in many routine clinical data sets, there were unfortunately no therapist characteristics available for analysis.

2.2 | Sample

The CORE National Research Database 2011 (Stiles, Barkham, & Wheeler, 2015) provided the sample pool for this study. The database comprises routine clinical data from n = 104,474 patients seen by 2,442 therapists across 52 clinics in seven sectors across the United Kingdom. Sectors were as follows: primary care, secondary care, tertiary care, university counselling, workplace counselling, voluntary, and private practice. Ethical approval was covered by National Research Ethics Service Application No. 05/Q1206/128 (Amendment 3).

The primary care sector is usually the first point of contact for people experiencing common mental health problems. Primary care services include general practitioner (GP) surgeries, community centres, and Improving Access to Psychological Therapies (IAPTs) programme sites. The database used in the current study did not include IAPT sites, however, due to differences in outcome recording procedures.

Secondary care services are generally more specialized and treat more complex or severe mental health difficulties. Secondary care services can typically only be accessed via referral from another primary or secondary care service. Tertiary care services are even more specialized again. Alternatively, patients can access treatment via university or workplace counselling centres, voluntary organizations/charities, or through private care organizations.

Inclusion criteria for Patients were (i) aged 16–95, (ii) received individual therapy, (iii) had recorded employment and session attendance data, and (iv) had valid pre-therapy and post-therapy CORE-OM scores (≥90% items completed). Further analysis-specific inclusion criteria were then applied in order to ensure cluster sizes were robust (Schiefele et al., 2017), requiring that (i) each therapist had seen ≥10 patients and (ii) each sector had ≥10 eligible therapists (Figure 1).

The included sample comprised routine clinical data from N = 26,814 patients, seen by 466 therapists, across 40 clinics with numbers in each sector as follows: primary (9,106 patients, 102 therapists, and 5 clinics), secondary (995 patients, 27 therapists, and 6 clinics), university (5,472 patients, 75 therapists, and 10 clinics), voluntary (4,794 patients, 171 therapists, and 12 clinics), and workplace (6,447 patients, 91 therapists, and 7 clinics). Data from private (N = 287) and tertiary (N = 62) sectors were excluded by the final inclusion criterion.

Therapists saw a mean of 57.5 patients each (standard deviation [SD] = 68.8). Clinics had a mean of 11.7 therapists each (SD = 12.6), whereas sectors had a mean of 93.2 therapists each (SD = 52.1). Mean age for the sample was 38.2, with 69% female patients. According to the CORE-OM clinical threshold, 89% of patients met criteria for clinical distress.

Included patients (compared with patients excluded due to exclusion criteria) were more likely to be female, white, older, and employed, with lower pre-therapy and post-therapy CORE-OM scores. Included patients on average attended more sessions, had higher attendance, less frequent sessions, and were more likely to have a planned ending (all Bonferroni adjusted $p \leq .004$; Table 1).

2.3 | Analysis

Multilevel modelling (MLM) accounts for the hierarchical dependence inherent in provision of psychological interventions (i.e., each therapist working with multiple patients). MLM simultaneously models variance at all levels of the hierarchy (Raudenbush & Bryk, 2002), including random effects (Martindale, 1978).

Multilevel models used a two-stage estimation process, involving iterative generalized least squares (IGLS) estimation followed by Markov chain Monte Carlo (MCMC) estimation (Browne, 2016; Rasbash, Steele, Browne, & Goldstein, 2012). IGLS estimation is efficient to run and uses relatively automated procedures. As such, it is a common first-stage estimation approach. MCMC requires more time-consuming calculations and prior distributions to be input but is able to calculate Bayesian credible intervals for coefficients that can be robustly statistically compared. For this reason, it was employed as a second-stage estimation approach, using prior distributions generated by the IGLS estimation.

Multilevel models with two levels (patient and therapist) were developed for each sector. As such, the study used data from one sample, split into five subsamples, one for each sector. MLM requires sufficient cluster sizes at each level to allow robust and accurate estimation. Small cluster sizes (particularly at the top level) can lead to problems including negative bias in standard errors (Maas & Hox, 2004). A minimum of 50 top level clusters has been recommended by some (Maas & Hox, 2004), although resilient methods such as MCMC estimation exist to enable the analysis of smaller samples (McNeish & Stapleton, 2016). In this case, there were considered to be insufficient sectors (five) to be able to use sector as a clustering level. The use of a separate model for each sector also permits easily interpretable comparison of therapist effects between sectors and was therefore also a design choice. Again, although the potential for three-level models was explored, within these subsamples, there were too few clinics (average eight per sector) to achieve stable models including clinics as a separate level. As such, clinics were modelled as fixed effects in the models (i.e., using a categorical variable to represent the clinic).

The therapist effect was defined as the proportion of overall unexplained variance in clinical outcome attributable to the therapist

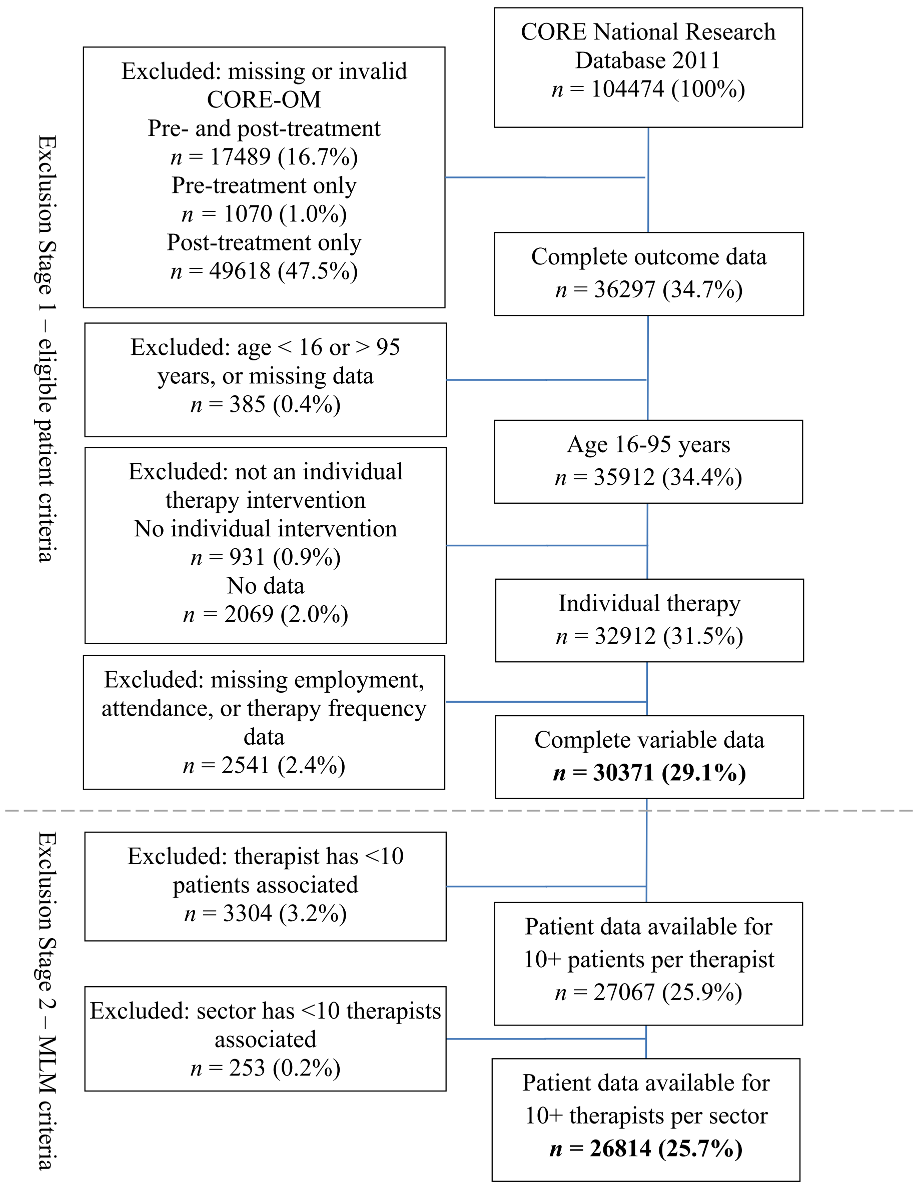

FIGURE 1 Inclusion flowchart. CORE-OM, Clinical Outcomes in Routine Evaluation-Outcome Measure; MLM, multilevel modelling [Colour figure can be viewed at wileyonlinelibrary.com]

level (the variance partition coefficient [VPC]). MCMC estimation was used to calculate 95% credible intervals. Multilevel models (like most analyses) make statistical assumptions, including that residuals are homoscedastic and normally distributed. In order to meet these assumptions, level 1 outcome variance was modelled as a function of nonrisk severity.

The significance of each control variable, interaction, and random effect was determined in two ways. First, the reduction in −2*log-likelihood score (indicating model fit) was required to exceed the corresponding chi-square critical value. Second, the Z score of the coefficient was also required to exceed the critical value for 95% confidence (1.96).

Models were tested in the following order. First, random intercepts were tested in a null model. Second, control variables were each tested. Where control variables were significant using linear coefficients, polynomial terms were also tested to assess whether they significantly improved model fit. Third, random slopes were tested for each significant control variable. Fourth, interactions between significant control variables were tested.

Continuous variables were patient post-therapy severity score, pre-therapy nonrisk score, pre-therapy risk score, age, number of sessions attended, and percentage of sessions attended. Categorical variables were patient employment status (employed, other role, and not employed), therapy frequency (more than weekly, weekly, less than weekly, and no fixed frequency), and clinic (each clinic was coded as a separate category).

Employment status categories were derived using Wald tests and inspection of coefficients before the main analysis from a broader range of initial categories: *not employed* (comprising *receiving benefits*, *unemployed*, and *retired*), *other role* (comprising *part-time student*, *full-time student*, *houseperson*, *other*, and *N/A*), and *employed* (comprising *part-time employment* and *full-time employment*).

Wald tests were used within the main model analyses to compare categories and combine those whose coefficients did not differ

TABLE 1 Comparisons between included and excluded patients

Variable	Included sample	Excluded patient	Included/Excluded difference
Patients	26,814	77,660	
Mean patients per therapist (SD)	57.5 (68.8)		
Mean therapists per clinic (SD)	11.7 (12.6)		
Mean patient age (SD)	38.2 (13.0)	34.9 (13.2)[a]	$t(99{,}309) = 35.0, p < .001$
Female	69%	66%[b]	$\chi^2(1{,}N = 103{,}082) = 82.39, p < .001$
White	88%[c]	81%[d]	$\chi^2(1{,}N = 92{,}544) = 599.353, p < .001$
Employment status			
In work	58%	41%[e]	$\chi^2(2{,}N = 100{,}084) = 2411.022, p < .001$
Other role	27%	41%[e]	
Not employed	14%	18%[e]	
Mean pre-therapy CORE (SD)	17.9 (6.3)	18.2 (6.8)[f]	$t(55{,}870.287) = -7.48, p < .001$
Mean pre-therapy nonrisk CORE (SD)	20.9 (6.9)	21.2 (7.4)[g]	$t(54{,}611.427) = -5.75, p < .001$
Mean pre-therapy risk CORE (SD)	3.9 (5.7)	4.8 (6.6)[h]	$t(59{,}210.991) = -20.86, p < .001$
Mean post-therapy CORE (SD)	8.9 (6.4)	9.6 (6.9)[i]	$t(18{,}105.580) = -9.48, p < .001$
Mean therapy sessions (SD)	8.3 (9.4)	7.6 (13.7)[j]	$t(62378.481) = 7.24, p < .001$
Mean session attendance % (SD)	90% (16%)	76% (26%)[j]	$t(60{,}064.939) = 82.485, p < .001$
Therapy frequency			
More than weekly	1%	1%[k]	
Weekly	56%	64%[k]	$\chi^2(3{,}N = 55{,}114) = 684.423, p < .001$
Less than weekly	30%	21%[k]	
No fixed frequency	14%	14%[k]	
Planned ending %	92%[l]	50%[m]	$\chi^2(1{,}N = 63{,}543) = 12656.671, p < .001$

Abbreviations: CORE, Clinical Outcomes in Routine Evaluation; SD, standard deviation.
[a]N = 72,497.
[b]N = 76,268.
[c]N = 26,240.
[d]N = 66,304.
[e]N = 73,270.
[f]N = 59,101.
[g]N = 61,230.
[h]N = 61,150.
[i]N = 10,553.
[j]N = 35,970.
[k]N = 28,300.
[l]N = 26,717.
[m]N = 36,826.

significantly from each other. Data were analysed using IBM SPSS Statistics and MLwiN (Rasbash, Charlton, Browne, Healy, & Cameron, 2016).

3 | RESULTS

3.1 | Multilevel models

Descriptive demographic statistics across sectors is shown in Table 2. Summary multilevel model specifications are shown in Table 3 (full specifications are included in the supporting information). Significant random intercepts were detected in all five sectors, indicating the presence of therapist effects. Nonrisk severity, risk-related severity, sessions attended, and clinic attended were significant in all sectors. Nonrisk severity and risk-related severity were associated with poorer outcome. Attending more sessions was also associated with poorer outcome in all but the voluntary sector. In contrast, percentage of sessions attended was significant in all but the secondary care sector, and was consistently associated with more positive outcome. Age was significant in all but the university sector, such that older patients, experienced poorer outcome. Employment status was significant in all but the university and workplace sectors - being employed were associated with the most positive outcomes, followed by patients in other roles, with patients who were not employed experiencing poorest outcomes. Therapy frequency was not significant in any sector. A number of polynomial coefficients and interactions were detected (full model specifications available as supporting information). A number of

TABLE 2 Comparison of intake characteristics between sectors

Variable	Primary care	Secondary care	University	Voluntary	Workplace
Patients	9,106	995	5,472	4,794	6,447
Mean patient age (SD)	42.0 (13.6)	41.4 (13.4)	26.2 (8.7)	38.2 (10.8)	42.5 (10.2)
Employment status[a]					
In work (%)	59	42	11	63	96
Other role (%)	17	11	88	13	3
Not employed (%)	24	47	1	24	1
Mean pre-therapy CORE (SD)	18.4 (6.2)	20.6 (7.0)	17.6 (6.1)	17.3 (6.5)	17.3 (6.1)
Mean pre-therapy nonrisk CORE (SD)	21.5 (6.8)	23.4 (7.3)	20.5 (6.7)	20.1 (7.2)	20.4 (6.8)
Mean pre-therapy risk CORE (SD)	4.1 (5.8)	7.6 (8.0)	3.9 (5.6)	4.2 (5.7)	2.8 (4.7)

Abbreviations: CORE, Clinical Outcomes in Routine Evaluation; SD, standard deviation.
[a]Employment status categories defined as follows: employed (part-time employment and full-time employment), not employed (receiving benefits, unemployed, and retired), and other role (part-time student, N/A, houseperson, full-time student, and other).

TABLE 3 Summary multilevel model specifications by sector

Specification	Primary	Secondary	University	Voluntary	Workplace
Constant	9.723^{R}	13.541^{R}	8.935^{R}	9.784^{R}	7.979^{R}
Nonrisk severity (polynomial ^{2} term)	0.328^{RX}	0.461^{X}	0.302^{X}	0.309^{X}	0.268^{X}
	—	—	—	0.003	—
Risk severity (polynomial ^{2} term)	0.104	0.132	0.108^{RX}	0.165^{RX}	0.222
	—	0.009	—	—	−0.007
Age (polynomial ^{2} term)	0.029^{X}	0.024	-	0.027^{X}	0.028^{X}
	—	—	—	—	0.001
Employed	−1.400	−1.073	—	−1.377	—
Other role	−0.817	Reference	—	−0.626	—
Not employed	Reference	Reference	—	Reference	—
Sessions (polynomial ^{2} term)	0.184^{R}	0.043^{X}	0.049^{X}	−0.033	0.108^{R}
	0.001	—	—	0.0003	−0.001
Attendance (polynomial ^{2} term)	−0.033^{X}	—	−0.033^{X}	−0.017^{RX}	−0.027
	—	—	—	—	−0.0003
Patients	9,106	995	5,472	4,794	6,447
Therapists	102	27	75	171	91
Clinics	5	6	10	12	7
Mean patients per therapist (SD)	89.3 (90.7)	36.9 (27.9)	73.0 (60.5)	28.0 (29.5)	70.8 (84.1)
Mean therapists per clinic (SD)	20.4 (15.8)	4.5 (4.8)	7.5 (6.8)	14.3 (18.0)	13.0 (5.8)

Note. Variable coefficients are for linear effects, unless stated otherwise (polynomial square terms are noted where significant).^{R} denotes significant random effect, and ^{x} denotes significant interaction with another variable.

interactions were detected, most notably involving pre-therapy severity, attendance percentage, and sessions attended.

3.2 | Therapist effects

MCMC-estimated therapist effects after controlling for variables and clinics were as follows (95% credible intervals in parentheses): primary care = 8.4% (5.8%–11.7%), secondary care = 4.1% (0.9%–9.3%), university = 2.1% (1.1%–3.7%), voluntary = 2.3% (0.4%–2.2%), and workplace = 1.1% (0.4%–2.2%). These results are shown in Figure 2.

Directly comparing overlap between credible intervals can produce conservatively biased assessments of significance. As such, 95% credible intervals for the between-sector *difference* in MCMC-estimated therapist effect chains were used to compare the primary care sector with each other sector. The primary care therapist effect

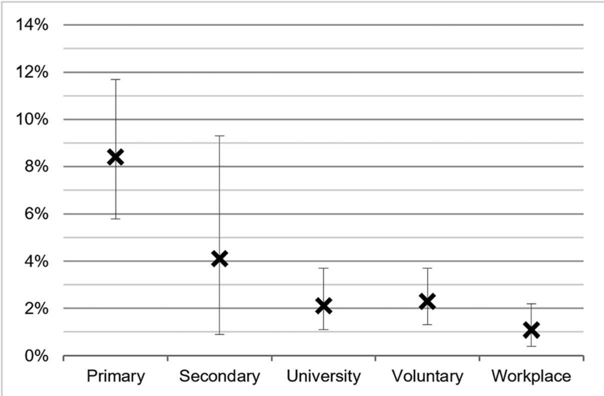

FIGURE 2 Therapist effect estimates and 95% credible intervals for each sector, after controlling for variance at the clinic level

(8.4%) was significantly larger than the therapist effect in every other sector (1.1%–2.3%) except for secondary care (4.1%). There were no significant differences between the secondary, university, voluntary, and workplace sectors.

Significant random slopes were detected, indicating that therapist effects were moderated by measured variables. In the primary care sector, the therapist effect was larger for patients with higher nonrisk severity scores. In the university and voluntary sectors, the therapist effect was larger for patients with higher risk severity scores. In the primary and workplace sectors, the therapist effect was larger for patients receiving higher numbers of sessions. Finally, in the voluntary sector, the therapist effect was larger for patients with fewer missed sessions.

4 | DISCUSSION

Finding that the therapist effect in the primary care sector was substantially greater than in the university, voluntary, and workplace sectors (which were comparable with each other) underlines the importance of considering the context of care in understanding therapist effects. The therapist effect in the secondary care sector was intermediate, with no significant differences from any other sector. In the United Kingdom, the National Health Service (NHS) provides the therapy in both primary care and secondary care sectors but not in the other sectors. The observed differences did not appear to be explained by patient intake severity, age, employment status, or consistent attendance. Nor were they explained by treatment duration or the clinic where patients received treatment. Notably, the non-NHS sector therapist effects were qualitatively lower than is typically reported in practice-based studies (Baldwin & Imel, 2013; Johns et al., 2019). This is consistent with Johns et al.'s (2019) findings of lower therapist effects in university specific practice-based studies.

It is difficult to do more than speculate about mechanisms underlying our findings, particularly given the broad credible intervals for secondary care. One hypothesis is that therapists in primary care settings (or in NHS settings more generally) tend to be relatively more variable in their skill or practice characteristics compared with other non-NHS care contexts. This is arguably unintuitive, given the NHS is a national, single parent organization with an embedded system of standards and policies, compared with the multiple universities, voluntary organizations, or workplaces, each with potentially different policies, funding streams, therapist recruitment practices, codes of professional practice, and so forth. It may be helpful to examine other factors that may have contributed to increased therapist variability in the primary care sector, such as variation in training/experience working with the relevant population, or the effects of burnout.

A second hypothesis may be that differences in therapist effects are partially accounted for by differences in patient population or organizational structure. For example, non-NHS clinics may exhibit relatively greater heterogeneity in nontherapist factors (e.g., organizational factors), thereby reducing the relative contribution of the therapist to outcome. Clinical populations in some care contexts may be much more highly selected than in other contexts. In principle, this might *reduce* variance at the *patient* level, thereby potentially *increasing* the relative therapist effect. However, statistical associations may be more complex; for example, a homogenous high-severity population might be expected to have greater therapist effects than a population with more heterogeneous (and therefore lower average) severity (Saxon & Barkham, 2012). Potential organizational factors include constraints on treatment duration (e.g., imposed by public funding and political pressures). It has been argued that therapeutic dyads tend to regulate treatment duration to optimize outcomes (Stiles et al., 2015). If duration is relatively free to vary, therapists who work less quickly may have longer treatments but similar mean outcomes, reducing therapist effects. The current study controlled for treatment duration, so this specific factor is less likely to account for observed differences. However, there may be other contributory organizational factors that were unavailable in the current study. As such, it is difficult to make specific conclusions from the current findings alone.

A third hypothesis is that the differences could reflect undetected neighbourhood effects (Firth et al., 2019). Therapists in primary care tend to work in locations more embedded in communities than do therapists in centralized clinics. Systematic differences in neighbourhood deprivation (or other neighbourhood/geographic variables) between therapists' caseloads may impact differentially on therapists' outcomes. Without explicit modelling of a neighbourhood level, this variance may be inappropriately assigned to the therapist level. This study did control for clinic effects, but it was not possible to appropriately model neighbourhood effects (e.g., using a three-level model), due to sample constraints. Despite this, the study addressed a number of methodological factors identified by Johns et al. (2019) as potential confounders. These included using a single practice-based data set, a single well-supported analytical approach, a single reliable and validated outcome measure, and sufficient sample size. Because inclusion criteria required pre- and post-therapy symptom severity scores and noncompleters rarely completed the post-therapy forms, patients who did not complete treatment were much less likely to be included. Thus, it cannot be assumed that the findings of the current study will generalize to patients who do not complete treatment.

The finding that therapist effects may vary according to context should be considered in research design, clinical application of research evidence, treatment planning, and psychological care provision. Analyses should control for relevant sources of variability such as therapist effects, clinic effects, and neighbourhood effects. Research cannot be assumed to generalize beyond specific care contexts. For example, research findings from a study conducted in a university clinic may underestimate the variability between therapists in other care contexts and therefore overestimate differential treatment effects in those contexts. Further research and routine evaluation should be undertaken to understand how therapist effects may vary in specific clinical delivery contexts (e.g., international/cultural differences and specific clinical populations).

In clinical practice, a care context with larger therapist effects might focus more resources on understanding differences in effectiveness between their therapists; for example, by seeking to (i) understand potential confounding systematic differences in caseload or care provision, (ii) improve the outcomes of less effective therapists (e.g., using techniques such as deliberate practice; Chow et al., 2015), or (iii) systematize what is working well for more effective therapists. In contrast, a care context with smaller therapist effects might focus on more generalized initiatives (i.e., those applicable to all therapists) or seek to understand and act on variability in other parts of the care system (either intraclinic or interclinic—e.g., standardization of treatment, accessible appointments, and waiting lists). Very low therapist effects might reflect a limitation or bottleneck in the care system that is preventing therapists from fully contributing to outcome (e.g., because of very high dropout rates). Of course, any interpretation of therapist effects or resultant actions also depend on other clinical information, such as average outcomes in the service; for example, in the case of low therapist effects, are therapists consistently enabling recovery, or consistently failing to achieve clinical change?

5 | CONCLUSIONS

Variance in therapists' outcomes appeared far greater (up to 4–8×) in NHS primary care than in non-NHS sectors. Estimates of the size of therapist effects should be understood in relation to the context from which they were derived rather than general characteristics. This may have implications in the design and application of research evidence, treatment planning, and the delivery of psychological care provision. Differences in state versus private provision, or front-line versus specialist provision may be useful to explore in future research.

ACKNOWLEDGEMENTS

We would like to thank John Mellor-Clark and Alex Curtis Jenkins at CORE-IMS for their assistance in collating the data set. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

MB was a developer of the CORE measures.

ORCID

Nick Firth https://orcid.org/0000-0003-1984-6869

REFERENCES

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). New York, NY: Wiley.

Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry*, *186*, 239–246. https://doi.org/10.1192/bjp.186.3.239

Barkham, M., Lutz, W., Lambert, M. J., & Saxon, D. (2017). Therapist effects, effective therapists, and the law of variability. In L. G. Castonguay, & C. E. Hill (Eds.), *Therapist effects: Toward understanding how and why some therapists are better than others* (pp. 13–36). Washington, DC: American Psychological Association.

Barkham, M., Mellor-Clark, J., Connell, J., Evans, R., Evans, C., & Margison, F. (2010). The CORE measures & CORE system: Measuring, monitoring, and managing quality evaluation in the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 175–219). Chichester: Wiley.

Barkham, M., Mullin, T., Leach, C., Stiles, W. B., & Lucock, M. (2007). Stability of the CORE-OM and the BDI-I prior to therapy: Evidence from routine practice. *Psychology and Psychotherapy: Theory, Research and Practice*, *80*, 269–278. https://doi.org/10.1348/147608306x148048

Bohart, A. C., & Greaves Wade, A. (2013). The client in psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.) (pp. 219–257). New York, NY: Wiley.

Browne, W. J. (2016). MCMC estimation in MLwiN: Centre for multilevel modelling, University of Bristol.

Cahill, J., Barkham, M., Stiles, W. B., Twigg, E., Hardy, G. E., Rees, A., & Evans, C. (2006). Convergent validity of the CORE measures with measures of depression for clients in cognitive therapy for depression. *Journal of Counseling Psychology*, *53*, 253–259. https://doi.org/10.1037/0022-0167.53.2.253

Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy*, *52*, 337–345. https://doi.org/10.1037/pst0000015

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O., & Miles, J. N. V. (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *British Journal of Psychiatry*, *190*, 69–74. https://doi.org/10.1192/bjp.bp.105.017657

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, *180*, 51–60. https://doi.org/10.1192/bjp.180.1.51

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

Hamilton, K. E., & Dobson, K. S. (2002). Cognitive therapy of depression: Pretreatment patient predictors of outcome. *Clinical Psychology Review*, *22*, 875–893. https://doi.org/10.1016/s0272-7358(02)00106-x

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

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*, 127–137. https://doi.org/10.1046/j.0039-0402.2003.00252.x

Martindale, C. (1978). Therapist-as-fixed-effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *46*, 1526–1530. https://doi.org/10.1037//0022-006x.46.6.1526

McNeish, D., & Stapleton, L. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, *51*, 495–518. https://doi.org/10.1080/00273171.2016.1167008

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2016). MLwiN (Version 2.36): Centre for multilevel modelling, University of Bristol.

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2012). A user's guide to MLwiN.

Raudenbush, S. W., & Bryk, A. S. (Eds.) (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

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., … 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

Stiles, W. B., Barkham, M., & Wheeler, S. (2015). Duration of psychological therapy: Relation to recovery and improvement rates in UK routine practice. *British Journal of Psychiatry*, *207*, 115–122. https://doi.org/10.1192/bjp.bp.114.145565

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Firth N, Saxon D, Stiles WB, Barkham M. Therapist effects vary significantly across psychological treatment care sectors. *Clin Psychol Psychother*. 2020;1–9. https://doi.org/10.1002/cpp.2461