Data supplement to Delgadillo et al. Impact of long-term medical conditions on the outcomes of psychological therapy for depression and anxiety. *Br J Psychiatry* doi: 10.1192/bjp.bp.116.189027

Online supplement DS1

Differential item functioning (DIF) analysis

The following eight tables present the results of the analysis for differential item functioning (DIF) in more detail. Details on the steps of the procedure will be presented using the example of the first item of the PHQ-9 (PHQ1) at the first available assessment (tables DS1 and DS2). All models were estimated in Stata 14 (StataCorp, College Station, Texas, 2015).

For unidimensional and unbiased instruments it would be expected that the response of a single item can be predicted to a large extent with an estimate of the latent variable that the instrument is supposed to measure. DIF would be deemed to present in a specific item, if after controlling for the latent variable other variables would still be predictive of the item response. Approaches based on logistic regression models[17] operationalise this straight forwardly: In a first step, the item to be tested is regressed upon an estimate of the latent trait and in a second step the variable is added that is tested for potential bias. In our example, the estimate of the latent trait when predicting PHQ1 was generated with a Generalised Partial Credit Model[34] for items PHQ2-PHQ9. In the first step (columns "Trait" in all tables), only this estimate was used to predict the response to item PHQ9 with an ordinal logistic regression model. The tables present the Pseudo-R^2 (e.g., .242 for PHQ1, table DS1) as well as the result of the comparison between this model and the baseline model without any predictors (Wald-$\chi^2_{df=1}$ = 12630.59; p<.001; table DS1). This comparison is highly significant and a relevant amount of variance in PHQ1 responses is explained, which indicates that items and trait are highly correlated as to be expected. This result is found for all items (across all tables).

In the second step of this approach, the grouping variable is added (columns "Trait + LTC" in all tables), in our case "no LTC vs. any LTC" (table DS1). The comparison test compares now the model including two predictors with the model with only the estimate of the latent variable (Wald-$\chi^2_{df=1}$ = 11.82; p<.001; table DS1). And while this test is highly significant (suggesting the potential of DIF), the main interpretation for the relevance of DIF is the comparison of the Pseudo-R^2 values of the two models. Generally a cut-off of >0.035 is judged to be indicative of DIF that could potentially affect group comparisons[17]. In this case, as well as for all other comparisons presented between these two models the differences in Pseudo-R^2 are well below this cut off, suggesting that the significant effects found in some comparisons are due to the statistical power of the large sample size.

The two steps so far have tested whether there is a constant difference in the probability of endorsing certain symptoms, in this case whether for example the group with any LTC shows, after controlling for distress, a higher or lower probability across the whole severity spectrum. This is usually called "uniform DIF".[35] But one could also imagine the case that the severity and LTC presence interact, e.g., that patients with higher levels of distress and an LTC are more likely to respond in the top categories of PHQ1. Such an effect would be called "non-uniform" DIF[35] and is tested for by adding in a third model the interaction term between the latent variable estimate and the grouping variable (columns "Trait X LTC" in all tables). The model test compares now the model including the interaction term with the model that contains only the latent variable estimate and the grouping variable. In the chosen example, this test is not significant (Wald-$\chi^2_{df=1}$ = 3.26; p=.07; table DS1), which indicates that the interaction term does not add any information above and beyond the independent effects of the two variables. Again, the relevant result would be if a difference in Pseudo-R^2 values would be detected

between this model and either the "Trait" or the "Trait + LTC" models, which is not the case for any of the tested items.

Per outcome measure (PHQ-9 and GAD-7) and assessment (first and last available) two tables are presented each: the first of these shows the results for no "LTC vs any LTC", while the second one presents the analyses with the LTC categories as used in the regression analysis presented in the main body of the research manuscript.

Overall, the increases in Pseudo-R^2 values for all items and all comparisons are very low, indicating that DIF is unlikely to have an impact on the results at either first or last assessments with both instruments

Table DS1 Results for DIF analyses of the PHQ-9 responses at the first recorded assessment comparing any LTC vs no LTC

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[a]	Trait X LTC[a]
PHQ1	.242	.243	.243	12630.59 (p<.001)	11.82 (p<.001)	3.26 (p=0.07)
PHQ2	.275	.275	.275	12699.73 (p<.001)	1.01 (p=0.32)	0.33 (p=0.57)
PHQ3	.136	.136	.136	7633.45 (p<.001)	23.23 (p<.001)	0.07 (p=0.79)
PHQ4	.170	.171	.171	8933.49 (p<.001)	54.45 (p<.001)	2.31 (p=0.13)
PHQ5	.133	.133	.133	8318.27 (p<.001)	5.87 (p=0.02)	0.45 (p=0.50)
PHQ6	.180	.180	.180	9906.78 (p<.001)	22.28 (p<.001)	0.09 (p=0.76)
PHQ7	.158	.158	.158	9677.69 (p<.001)	0.14 (p=0.71)	0.31 (p=0.58)
PHQ8	.106	.106	.106	6651.89 (p<.001)	13.84 (p<.001)	4.11 (p=0.04)
PHQ9	.121	.121	.121	6005.02 (p<.001)	9.36 (p=0.002)	0.25 (p=0.62)

Note. [a]Degrees of freedom = 1

Table DS2: Results for DIF analyses of the PHQ-9 responses at the first recorded assessment comparing all LTC groups (no LTC as reference category)

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[b]	Trait X LTC[b]
PHQ1	.242	.243	.243	12630.59 (p<.001)	48.10 (p<.001)	21.85 (p=0.01)
PHQ2	.275	.275	.275	12699.73 (p<.001)	7.48 (p=0.59)	15.45 (p=0.08)
PHQ3	.136	.137	.137	7633.45 (p<.001)	31.40 (p<.001)	5.26 (p=0.81)
PHQ4	.170	.171	.171	8933.49 (p<.001)	71.60 (p<.001)	13.35 (p=0.15)
PHQ5	.133	.133	.133	8318.27 (p<.001)	17.21 (p=0.05)	9.11 (p=0.43)
PHQ6	.180	.181	.181	9906.78 (p<.001)	61.89 (p<.001)	8.90 (p=0.45)
PHQ7	.158	.158	.158	9677.69 (p<.001)	15.97 (p=0.07)	17.58 (p=0.04)
PHQ8	.106	.106	.107	6651.89 (p<.001)	28.87 (p<.001)	9.83 (p=0.36)
PHQ9	.121	.121	.121	6005.02 (p<.001)	25.14 (p=0.003)	8.98 (p=0.44)

Note. [a]Degrees of freedom = 1; [b]Degrees of freedom = 9

Table DS3: Results for DIF analyses of the PHQ-9 responses at the last recorded assessment comparing any LTC vs no LTC

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[a]	Trait X LTC[a]
PHQ1	.407	.407	.407	14115.43 (p<.001)	1.24 (p=0.27)	8.74 (p=0.003)
PHQ2	.396	.396	.397	14438.18 (p<.001)	0.06 (p=0.81)	5.36 (p=0.02)
PHQ3	.252	.252	.253	12792.29 (p<.001)	26.68 (p<.001)	5.21 (p=0.02)
PHQ4	.298	.298	.298	13947.03 (p<.001)	38.67 (p<.001)	0.36 (p=0.55)
PHQ5	.247	.248	.248	11793.81 (p<.001)	6.03 (p=0.01)	3.97 (p=0.05)
PHQ6	.317	.317	.317	13498.00 (p<.001)	16.91 (p<.001)	0.01 (p=0.94)
PHQ7	.306	.306	.306	13049.03 (p<.001)	0.21 (p=0.65)	2.88 (p=0.09)
PHQ8	.244	.244	.244	9353.13 (p<.001)	9.42 (p=0.002)	0.00 (p=0.97)
PHQ9	.253	.253	.253	6238.94 (p<.001)	2.70 (p=0.10)	0.02 (p=0.89)

Note. [a]Degrees of freedom = 1

Table DS4: Results for DIF analyses of the PHQ-9 responses at the last recorded assessment comparing all LTC groups (no LTC as reference category)

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[b]	Trait X LTC[b]
PHQ1	.407	.407	.407	14115.43 (p<.001)	27.04 (p=0.001)	10.36 (p=0.32)
PHQ2	.396	.397	.397	14438.18 (p<.001)	12.14 (p=0.21)	8.59 (p=0.48)
PHQ3	.252	.253	.253	12792.29 (p<.001)	36.98 (p<.001)	12.32 (p=0.20)
PHQ4	.298	.298	.298	13947.03 (p<.001)	49.56 (p<.001)	9.67 (p=0.38)
PHQ5	.247	.248	.248	11793.81 (p<.001)	16.95 (p=0.05)	18.08 (p=0.03)
PHQ6	.317	.318	.318	13498.00 (p<.001)	39.79 (p<.001)	11.34 (p=0.25)
PHQ7	.306	.306	.306	13049.03 (p<.001)	17.23 (p=0.04)	7.31 (p=0.61)
PHQ8	.244	.245	.245	9353.13 (p<.001)	31.10 (p<.001)	22.32 (p=0.01)
PHQ9	.253	.253	.254	6238.94 (p<.001)	8.90 (p=0.45)	10.83 (p=0.29)

Note. [a]Degrees of freedom = 1; [b]Degrees of freedom = 9

Table DS5: Results for DIF analyses of the GAD-7 responses at the first recorded assessment comparing any LTC vs no LTC

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[a]	Trait X LTC[a]
GAD1	.238	.238	.238	11000.50 (p<.001)	0.14 (p=0.71)	3.19 (p=0.07)
GAD2	.332	.332	.333	12639.41 (p<.001)	0.40 (p=0.53)	12.68 (p=0.0004)
GAD3	.329	.329	.329	12306.44 (p<.001)	1.22 (p=0.27)	0.09 (p=0.76)
GAD4	.218	.218	.218	11352.83 (p<.001)	15.96 (p<.001)	0.60 (p=0.44)
GAD5	.118	.118	.118	7712.67 (p<.001)	20.48 (p<.001)	1.32 (p=0.25)
GAD6	.088	.088	.088	5840.43 (p<.001)	7.26 (p=0.007)	4.74 (p=0.03)
GAD7	.148	.148	.148	9292.46 (p<.001)	7.37 (p=0.007)	1.93 (p=0.16)

Note. [a]Degrees of freedom = 1

Table DS6: Results for DIF analyses of the GAD-7 responses at the first recorded assessment comparing all LTC groups (no LTC as reference category)

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[b]	Trait X LTC[b]
GAD1	.238	.238	.238	11000.50 (p<.001)	9.25 (p=0.41)	9.77 (p=0.37)
GAD2	.332	.333	.333	12639.41 (p<.001)	6.10 (p=0.73)	24.19 (p=0.004)
GAD3	.329	.330	.330	12306.44 (p<.001)	16.75 (p=0.05)	3.78 (p=0.93)
GAD4	.218	.219	.219	11352.83 (p<.001)	31.71 (p<.001)	4.64 (p=0.86)
GAD5	.118	.119	.119	7712.67 (p<.001)	44.53 (p<.001)	9.15 (p=0.42)
GAD6	.088	.089	.089	5840.43 (p<.001)	35.15 (p<.001)	10.70 (p=0.30)
GAD7	.148	.148	.148	9292.46 (p<.001)	24.97 (p=0.003)	6.27 (p=0.71)

Note. [a]Degrees of freedom = 1; [b]Degrees of freedom = 9

Table DS7: Results for DIF analyses of the GAD-7 responses at the last recorded assessment comparing any LTC vs. no LTC

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[a]	Trait X LTC[a]
GAD1	.384	.384	.384	14742.80 (p<.001)	0.79 (p=0.37)	2.89 (p=0.09)
GAD2	.461	.461	.461	14700.92 (p<.001)	2.35 (p=0.13)	1.36 (p=0.24)
GAD3	.455	.455	.455	14787.63 (p<.001)	0.66 (p=0.42)	0.92 (p=0.34)
GAD4	.343	.343	.344	14419.29 (p<.001)	3.42 (p=0.06)	2.38 (p=0.12)
GAD5	.260	.260	.260	11245.83 (p<.001)	15.78 (p<.001)	6.72 (p=0.01)
GAD6	.234	.235	.235	12058.38 (p<.001)	1.36 (p=0.24)	2.18 (p=0.14)
GAD7	.288	.288	.288	12396.65 (p<.001)	1.48 (p=0.22)	0.00 (p=0.95)

Note. [a]Degrees of freedom = 1

Table DS8: Results for DIF analyses of the GAD-7 responses at the last recorded assessment comparing all LTC groups (no LTC as reference category)

	Pseudo-R^2			Wald-χ^2 (p-value)		
	Trait	Trait + LTC	Trait X LTC	Trait[a]	Trait + LTC[b]	Trait X LTC[b]
GAD1	.384	.384	.384	14742.80 (p < .001)	15.70 (p=0.07)	11.12 (p=0.27)
GAD2	.461	.461	.461	14700.92 (p < .001)	12.24 (p=0.20)	18.43 (p=0.03)
GAD3	.455	.455	.455	14787.63 (p < .001)	4.39 (p=0.88)	4.14 (p=0.90)
GAD4	.343	.344	.344	14419.29 (p < .001)	34.03 (p<0.001)	9.38 (p=0.40)
GAD5	.260	.261	.261	11245.83 (p < .001)	39.82 (p<.001)	15.66 (p=0.07)
GAD6	.234	.235	.235	12058.38 (p < .001)	32.08 (p<.001)	10.45 (p=0.32)
GAD7	.288	.288	.288	12396.65 (p < .001)	9.33 (p=0.41)	13.37 (p=0.15)

Note. [a]Degrees of freedom = 1; [b]Degrees of freedom = 9

Online supplement DS2

Assessing the impact of LTC on post-treatment depression (PHQ-9) and anxiety (GAD-7): Low versus high intensity treatment pathways

To test for potential interaction effects of the intensity of treatment provided and the LTCs, we re-ran the model described in table 3 in the main body of this manuscript separately for cases that finished their treatment pathway after low intensity interventions (n = 18 902) and high intensity interventions (n = 8884). The following describes only differences that were found in significance or direction compared to the main analysis (detailed results in table DS9). Focusing only on the coefficients for the LTCs, we found very few changes compared to the analysis on the full sample. Both, COPD and Diabetes were only correlated with higher PHQ-9 (and for COPD: GAD-7) scores in patients finishing their treatment pathway after high intensity interventions. All other relationships remained the same.

Table DS9: Estimated LTC coefficients for the SUR model jointly predicting post-treatment depression (PHQ-9) and anxiety (GAD-7) severity for low and high intensity treatment pathways

LTC	Low intensity pathway		High intensity pathway	
	PHQ-9 (SE)	GAD-7 (SE)	PHQ-9 (SE)	GAD-7 (SE)
Asthma	0.18	0.26	0.16	0.10
	(0.17)	(0.15)	(0.24)	(0.22)
Cancer	-0.00	-0.46	0.88	0.45
	(0.64)	(0.57)	(0.99)	(0.88)
Chronic Musculoskeletal	1.84***	1.24***	1.01*	0.86*
	(0.32)	(0.28)	(0.46)	(0.40)
COPD	0.90	0.75	3.81***	2.34**
	(0.60)	(0.53)	(0.93)	(0.83)
Cardiovascular	0.05	-0.09	0.23	-0.00
	(0.26)	(0.23)	(0.43)	(0.38)
Diabetes	0.43	0.00	1.30**	0.50
	(0.34)	(0.30)	(0.51)	(0.45)
Epilepsy	-0.19	-0.05	0.25	0.33
	(0.50)	(0.44)	(0.80)	(0.71)
Severe Mental (psychotic)	2.93***	2.08***	4.13***	3.55***
disorder	(0.61)	(0.54)	(0.89)	(0.79)
Other LTC	0.65***	0.57***	0.67**	0.33
	(0.15)	(0.13)	(0.23)	(0.20)
Constant	4.31***	3.70***	4.82***	4.22***
	(0.32)	(0.28)	(0.49)	(0.44)
Observations	18 902	18 902	8884	8884
R-squared	0.38	0.33	0.31	0.27

Notes: SUR = seemingly unrelated regression; B = regression coefficient; SE = standard error; PHQ-9 = depression measure; GAD-7 = anxiety measure; LTC = long term medical condition; SE = standard error; COPD = chronic obstructive pulmonary disease; the reference category in this analysis = no self-reported long term condition; all coefficients controlled for demographic and treatment-related variables as for the main analysis (table 3); * p <.05; ** p <.01; *** p <.001

Additional references

34. Muraki E. A generalized partial credit model. In *Handbook of modern item response theory* (eds WJ van der Linden, RK Hambleton): 153-164. New York: Springer, 1997.

35. Teresi JA, Ocepek-Welikson K, Kleinman M, Eimicke JP, Crane PK, Jones RN, Lai JS, Choi SW, Hays RD, Reeve BB, Reise SP, Pilkonis PA, Cella D. Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. Psychol Sci Q 2009; **51**:148-180.