Multiverse to inform neurological research: an example using recovery outcome of neglect.

Margaret J Moore¹ and Nele Demeyere¹

¹University of Oxford, Department of Experimental Psychology, Radcliffe Observatory Quarter, Oxford, OX2 6GG, United Kingdom

Corresponding Author:

Prof. Nele Demeyere

nele.demeyere@psy.ox.ac.uk

+44 (0)1865 271 340

Author Contributions:

MJM conducted analysis and drafted the manuscript. ND conceptualised and supervised the study, acquired funding and edited manuscript drafts.

Wordcount: 3493

References: 40

Abstract: 249

Abstract:

Objective: Multiverse analysis provides an ideal tool for understanding how inherent, yet ultimately arbitrary methodological choices impact the conclusions of individual studies. With this investigation we aimed to demonstrate the utility of multiverse analysis for evaluating generalisability and identifying potential sources of bias within studies employing neurological populations.

Method: Multiverse analysis was used to evaluate the robustness of the relationship between post-stroke visuospatial neglect and poor long term recovery outcome within a sample of 1113 (age =72.5, 45.1% female) stroke survivors. A total of 25,600 t-test comparisons were run across 400 different patient groups defined using various combinations of valid inclusion criteria based on lesion location, stroke type, assessment time, neglect impairment definition, and scoring criteria across 16 standardised outcome measures.

Results: Overall, 33.9% of conducted comparisons yielded significant results. 99.9% of these significant results fell below the null specification curve, indicating a highly robust relationship between neglect and poor recovery outcome. However, the strength of this effect was not constant across all comparison groups. Comparisons which included <100 participants, pre-selected patients based on lesion type, or failed to account for allocentric neglect impairment were found to yield average effect sizes which differed substantially. Similarly, average effect sizes differed across various outcome measures with the strongest average effect in comparisons involving an activities of daily living measure and the weakest in comparisons employing a depression subscale.

Conclusions: This investigation demonstrates the utility of multiverse analysis techniques for evaluating effect robustness and identifying potential sources of bias within neurological research.

Keywords: methodology, behavioural neurology, multiverse, open science, stroke

Abbreviations:

OCS – Oxford Cognitive Screen

NIHSS – National Institute of Health Stroke Screen

SIS – Stroke Impact Scale

ADL – Activities of Daily Living

HADS – Hospital Anxiety and Depression Scale

Introduction:

Conducting any research study inevitably involves choosing specific analysis designs, outcome measures, and variables of interest from a multitude of possible choices. It is often unclear how these inherent, yet arbitrary methodological choices ultimately impact the final conclusions of any single study. It is plausible that the results of any one individual analysis can be drastically skewed, and therefore unrepresentative due to the specific combination of methodological choices made by the researchers[1]. This possibility is critically important to consider within the context of any neurological research.

Planning research within neurological populations requires researchers to select specific combinations of potentially valid methodological choices. For example, studies aiming to investigate post-stroke impairment must choose between various standard inclusion criteria with some studies opting to include patients with specific stroke types and including a sample selected regardless of stroke characteristics. Similarly, neurological investigations are required to choose from a wide range of equally valid behavioural outcome measures. For example, dementia-related cognitive impairment can be quantified using any one of multiple cognitive screening tools including the Mini Mental State Examination, Montreal Cognitive Assessment, or the ACE-R[2–4]. Neurological studies can also differ in inclusion criteria around the precise time window post diagnosis. For example, clinical trials investigating the efficacy of Parkinson’s treatment have commonly chosen to exclusively recruit patients within the early or later stages of disease progression[5]. These necessary choices are an inherent component of designing research within neurological populations, though they can sometimes be arbitrary (e.g. definitions of acute, subacute and chronic stages in stroke vary between studies[6]. However, the potential impact of these choices is not well understood.

In recent years, required pre-registration of protocols and trial registers have provided some degree of protection from spurious conclusions being drawn due to non-hypothesis

driven research[7]. However, these quality controls do not prevent biases arising from employing patient selection criteria and specific methodological choices. Regardless of the topic being investigated, methodological choices can ultimately impact the final conclusions of any single study. In order to better understand potential biases in selection measures, secondary analyses should be informed by a clear understanding of potential biases in selection measures which can be gleaned from secondary analyses of existing large data sets can help inform choices for strong confirmatory research designs.

Multiverse analysis provides an ideal methodology for evaluating the impact of potentially arbitrary design choices on the conclusions of neurological research[1,8]. In a multiverse analysis, a large number of potential datasets are generated by making each possible permutation of reasonable methodological choices within a single experiment. Each of these datasets are then independently analysed, and the cumulative results are used to evaluate the overall robustness of the effect of interest. Critically, multiverse analysis clarifies which specific methodological choices may lead to significant effects. Multiverse analysis is not yet a common approach in research within neurological populations, even though this analysis technique helps in evaluating the potential impact of inclusion criteria, variables of interest, and key outcome measures. This increase in transparency can potentially help evaluate past research where documented effects may have been partly or wholly due to the specific outcome measures employed, or more clearly identify potential limits in generalising results to wider patient populations. Given this method's potential for helping to evaluate the robustness of findings, more studies assessing the utility of multiverse analysis techniques within specific key questions addressed in neurological research are needed. Here, we provide an example multiverse analysis within post-stroke cognition research, specifically on the effect of acute visuo-spatial on longer term recovery outcomes, exploring the impact of decision making from our own study[9]

Demonstration of Multiverse Analysis:

Research investigating the predictive relationship between the occurrence of visuospatial neglect and poor long-term recovery outcome in stroke survivors provides an ideal context in which to investigate the utility of multiverse analysis within neurological research. Visuospatial neglect is a common cognitive deficit post-stroke, occurring in up to 48% of patients acutely[10], in which patients exhibit a marked inability to orient and attend to stimuli presented within one side of space[11]. A substantial body of evidence has found that the occurrence of acute visuospatial neglect acts as a strong predictor of poor recovery outcome in stroke survivor[12–14]. This association seems to be a particularly robust effect which is present within different neglect subtypes across multiple outcome measures, at multiple timepoints, and with varying inclusion criteria.

However, each investigation on this topic has made specific methodological choices which may potentially have impacted the study's results. For example, some previous investigations of the relationship between acute visuospatial neglect and poor outcome have included a sample restricted to only patients with right hemisphere damage (e.g. [14]). While visuospatial neglect commonly occurs following both right and left hemisphere lesions, neglect following right hemisphere damage is typically more severe and long lasting [12,15–16]. Therefore, including only right hemisphere patients in visuospatial neglect studies could potentially result in a sample which is biased towards more severe neglect cases and it is not yet clear whether these restricted samples are adequately representative of neglect within the full stroke population.

Similarly, previous studies have included samples of neglect patients recruited at different timepoints following stroke (e.g.[12,17]). It has been demonstrated that neglect follows a non-linear recovery trajectory throughout the first six months following stroke, with

most spontaneous recovery occurring within the first ten days following stroke[15]. This suggests that the “neglect” groups within more acute samples will include patients who will have improved by later timepoints and would therefore not be included in the “neglect” groups of studies who recruit patients at a later timepoint. It remains unclear whether and to what extent this difference in the specific patients included influenced the results of these studies.

Furthermore, visuospatial neglect is a highly heterogeneous condition with many different subtypes identified[18-19]. For example, some patients exhibiting impairment within an egocentric (self-centred) frame of reference and others within an allocentric (object-centred) reference frame [10, 20-21]. Egocentric and allocentric neglect represent doubly dissociated, independent cognitive impairments, though they also commonly co-occur within stroke survivors[10,22]. Many studies investigating the relationship between neglect and recovery outcome have exclusively included patients with egocentric neglect and have not included tests of allocentric neglect[13–15,17]. It therefore remains unclear whether studies which did not control for the potentially confounding impact of (additional or solely) allocentric neglect impairment produce externally valid conclusions.

Given that research investigating the predictive relationship between neglect and poor functional outcome in stroke survivors inherently requires such a wide range of methodological choices, this provides an ideal demonstration investigating the utility of multiverse analysis for evaluating the robustness and generalisability of effects within neurological research.

Methods:

This project represents a secondary analysis of longitudinal data collected as part of the Oxford Cognitive Screen (OCS) study (Demeyere et al 2015 - REC reference

11/WM/0299), OCS-Tablet Cognitive Screening study (REC reference 14/LO/0648) and OCS-Care study (Demeyere *et al.*, 2019 – REC reference 12/WM/00335). All 3 studies included acute cognitive screening and 6 month follow ups. All patients provided informed consent in line with the Declaration of Helsinki.

Participants:

Participants were included in this investigation if they had completed the OCS Cancellation Task during initial hospitalisation and had completed at least one outcome measure during either acute hospitalisation or follow-up.

Importantly, to accurately reflect the clinical reality in stroke, patients were not excluded based on stroke type, stroke side, or pre-existing neurological conditions, such as previous stroke. A total of 1159 patients (mean age = 72.5 years (SD= 13.7, range = 18-97), 45.1% female, average education = 11.9 years (SD= 3.82), 7.8% left handed) were included in this study. These participants had an average NIHSS score of 3 (SD=3.82 , range =0-22).

Stroke types of the final sample were reported as 909 ischemic, 173 haemorrhagic, 20 TIA, 6 unknown, and 51 not reported. Lesion sides were reported as 466 right, 415 left, 62 bilateral, 148 not visible, and 68 not reported. The average stroke-test interval was 8.1 days (SD=15.4) days. For the multiverse analysis, this participant group was divided into 200 subgroups, each selected based on a plausible combination of lesion location, stroke type, assessment time, and minimum score inclusion criteria (Table 1).

Inclusion Criteria

Impairment Definition

Lesion Location	Stroke Type	Assessment Time		Neglect Definition	Scoring Criteria
Right Hem.	Ischemic	≤7 d	< 2 M	Egocentric Only	Total Score <42 +
Left Hem.	Haemorrhagic	≤14 d	> 1 M	Allocentric Only	Asymmetry
Not Bilateral	Not a TIA	>7d & <1M	Known Time	Both Ego + Allo	Asymmetry Only
Any Visible	Any Known	< 1 M	Unrestricted	Any Neglect	
Unrestricted	Unrestricted				

Table 1: A visualisation of the inclusion criteria and impairment definition factors considered when creating multiverse subsamples with regards to the patient inclusion decisions. For each subsample, one option from each inclusion criteria category was selected, yielding 200 unique patient groups. Each of these 200 groups were then analysed for each of the possible methods for grouping patients with and without significant neglect (impairment definitions). Hem = Hemisphere, Ego = Egocentric, Allo = Allocentric.

Materials:

Each participant in this investigation completed the OCS Cancellation task[23] acutely and at least one standardised recovery outcome measure at 6 months follow-up. Here we considered measures of activities of daily life (including subscales), level of remaining stroke severity, measures of mood, and measures of quality of life. See Table 2 for a full list of all considered outcome measures. The OCS Hearts Cancellation Task has been demonstrated to be highly sensitive to neglect impairment (94.12%, versus the Behavioural Inattention Test Star Cancellation[23]) and to reliably differentiate between allocentric and egocentric visuospatial neglect deficits[22, 24-25].

The presence of egocentric neglect severity is scored by subtracting the number of correctly identified targets on the left side of the page from those correctly identified on the right side of the page[23] (Figure One). The cut-offs for egocentric neglect were asymmetry scores of less than -3 or greater than 3[23]. Allocentric neglect impairment on the OCS Cancellation task is scored by subtracting the total number of right-gap false positive responses from the number of left-gap false positives errors made. Cut-offs for impairment are allocentric scores of less than -1 or greater than 1. Some previous studies have employed a more conservative Hearts Cancellation Task Scoring method in which patients are required to have a significant egocentric asymmetry score and a total score of less than 42/50 in order to be classified as having egocentric neglect impairment[9-10]. This investigation employs this alternate scoring method and compares this method to the standard scoring approach (Table 1). For both egocentric and allocentric neglect, negative asymmetry scores denote right-lateralised neglect and positive asymmetry scores denote left-sided neglect.

Given that not all previous investigations have considered both egocentric and allocentric neglect impairment, four different methods for grouping participants into neglect/no neglect groups were employed (Table 1). In the first instance, egocentric impairment alone was considered as the “neglect” category while all other patients were categorised as “no neglect”. Next, allocentric neglect impairment alone was considered. In a third grouping, patients were required to exhibit both allocentric and egocentric neglect to be added to the “neglect” group. In the final instance, patients with any neglect impairment, regardless of the type, were grouped into the “neglect” group.

[Figure One]

Measure	Max Score	Time	N	Mean (SD)
SIS Total [26]	300	Follow-up	465	222.3 (46.3)
SIS ADL Total [26]	90	Follow-up	462	66.2 (17.5)
SIS Physical Weakness [26]	20	Follow-up	451	14.5 (4.5)
SIS Memory [26]	35	Follow-up	465	28.8 (6.2)
SIS Mood [26]	45	Follow-up	462	32.7 (6.4)
SIS Communication [26]	35	Follow-up	460	31.0 (5.0)
SIS Mobility [26]	45	Follow-up	464	34.2 (9.8)
SIS ADL[26]	45	Follow-up	462	39.2 (10.1)
SIS Affected Hand [26]	25	Follow-up	449	17.5 (6.9)
SIS ADL 2 [26]	45	Follow-up	457	27.4 (9.05)
NIHSS[27]	42	Acute	319	4.3 (3.8)
Quality of Life[28]	20	Follow-up	519	15.2 (3.1)
HADS Depression[29]	21*	Follow-up	491	5.5 (4.0)
HADS Anxiety[29]	21*	Follow-up	484	6.2 (3.9)
Carer Strain Index[30]	13*	Follow-up	145	7.6 (5.7)
Barthel Index[31]	100	Acute	517	17.4 (4.1)

Table 2: A list of the outcome measures considered within this investigation. N represents the number of participants for which data was available for each outcome measure. The Time column reports whether the assessment was delivered during acute hospitalization or at 6-month follow up. SIS = Stroke Impact Scale, NIHSS = National Institute of Health Stroke Scale, ADL = Activities of Daily Living, IADL = Instrumental Activities of Daily Living, HADS = Hospital Anxiety and Depression Scale. Max scores which are marked with a star denote tests in which a higher score represent worse outcome. These scores are inverted in all subsequent analyses. SIS total is the combined total of all SIS subtests. SIS ADL Total is the combined total of both SIS ADL measures.

Planned Analysis:

Multiverse analysis was employed to compare the level of impairment in patients with and without neglect. Multiverse analysis involves multiple permutations of any single base analysis (e.g. regression, t-test, ANOVA). Given this investigation aims to compare a series of continuous variables (e.g. outcome measures) between two categorical groups (neglect vs. no neglect), a t-test base analysis was employed. This base t-test analysis was performed for each possible permutation patient inclusion criteria (n=200), method of neglect impairment definition (n=8), and outcome measure (n=16) yielding 25,600 individual t-tests. Of these tests, only combinations leading to comparisons involving at least 5 neglect and 5 no-neglect participants were considered in subsequent analyses, resulting in a final sample of 18,898 valid comparisons. In line with multiverse methodology, and given that this analysis aims to simulate the findings of all potential individual analyses which could be conducted in isolation, no correction for multiple comparisons are employed[1,8].

First, a specification curve analysis[32] was conducted to evaluate the directionality and robustness of differences between participants with and without neglect. This analysis employed a null curve created by conducting all potential t-test analyses, but with participants randomly allocated to neglect/no neglect comparison groups. The overall generalisability of acute neglect's relationship with poor recovery outcome was analysed by first evaluating the overall proportion of conducted comparisons met traditional significance thresholds. Subsequent analyses aimed to identify sources of biases within these significant results by comparing average effect sizes (Cohen's d) across comparisons which employed different selection criteria. All data and code has either been made openly available on the Open Science Framework (<https://osf.io/vfnry/>) or is available upon request.

Results:

Specification Curve Analysis:

First, a specification curve analysis was conducted (Figure 1). Given the standard alpha level of 0.05, if no effect was present within the analysed data, approximately 5% of comparisons would be expected to yield significant results. However, 33.9% of the conducted comparisons were found to be statistically significant. These significant comparisons had an average effect size of -0.485 (SD=0.327, range=-3.87-0.552, 25th Quantile=-0.523, 75th Quantile=-0.334). This negative overall effect size indicates that as a whole, neglect was found to be associated with poorer performance across the various recovery outcome measures.

The robustness of this result was then evaluated by comparing this experimental specification curve to a null curve. The null curve representing randomly-allocated comparisons contained 3272/19200(17.0%) significant results with an average effect size of 0.163(SD=0.387, range =-1.23-1.12, 25th Quantile=-0.239, 75th Quantile=0.375). 99.9% of statistically significant experimental curve comparisons were found to fall below the null curve 25th quantile boundary (Figure 2). In line with interpretations of the null curve [32], this finding illustrates that the relationship between neglect and poor recovery outcome is highly robust and significant across many possible methodological combinations.

[Figure Two]

Evaluating the Impact of Methodological Decisions:

Within the robustness of the relationship between neglect and poor recovery outcome, nevertheless significant variance was present within the effect sizes of individual significant comparisons. For this reason, additional analyses were conducted to evaluate the impact of specific methodological choices on resulting effect sizes.

Given that the number of participants with available data within the possible different inclusion criteria choices, neglect grouping factor definitions and across the potential

outcome measures was found to vary dramatically, the impact of sample size on the proportion of significant results was explicitly addressed. Comparisons involving less than 100 participants were found to yield only 12.1% significant results while 46.5% of comparisons with 200-299 participants, and 100% of the comparisons including more than 500 participants yielded significant results. Figure 3 presents a visualisation of this known relationship between sample size and proportion of significant results in the present data.

[Figure Three]

Next, the impact of using different outcome measures was evaluated. Average effect sizes were found to vary across outcome measures ranging from -0.28 (HADS Depression) to -0.68 (NIHSS). Overall, comparisons involving HADS Depression ($M=-0.28$, $SD=0.08$), SIS IADL ($M=-0.32$, $SD=0.07$), SIS Mood ($M=-0.40$, $SD=0.16$) and Quality of Life ($M=-0.34$, $SD=0.09$) were found to result in the weakest effects and comparisons involving NIHSS ($M=-0.68$, $SD=0.29$), SIS ADL ($M=-0.63$, $SD=0.22$), SIS Total ($M=-0.57$, $SD=0.41$), and SIS Communication ($M=-0.55$, $SD=0.63$) were found to be the strongest (Figure 2, Figure 4).

[Figure Four]

The Impact of Significant Choice Combinations:

Figure 5 presents a visualisation of the impact of specific methodological choice combinations on comparison effect size. Given that many of the conducted comparisons involve small sample sizes which are associated with unreliable effect sizes[33-34], only comparisons involving samples of at least 108 participants were included in this comparison ($n=13,736$). This inclusion threshold was determined by calculating the minimum sample size needed to detect the average significant effect size (-0.485) with a power of 0.80.

Overall, the choice of outcome measure was found to have the greatest impact on comparison effect size across choice combinations with the average effect size ranging between -0.464 (SIS ADL) and -.0.113 (HADS Depression). Notably, these consistently negative average effect sizes demonstrate that acute neglect was associated with poor performance across all employed outcome measures. Other selection criteria led to similar ranges in average effect size. Comparisons which included all patients, regardless of lesion type resulted in stronger effects (mean effect=-0.310) than those which included patients with only right (-0.284) or left (-0.259) hemisphere damage (Figure 4). Similarly, analyses which considered allocentric neglect impairment generally resulted in stronger effects than those which considered only egocentric neglect impairment (-0.250). Comparisons which required participants to have significant asymmetry scores as well as total scores of <42 in order to be classified as exhibiting egocentric neglect (mean=-0.328) resulted in stronger average effect sizes than those which only required significant asymmetry scores (mean=-0.268). Assessment time and stroke type and selection criteria resulted in comparatively little variance within average effect sizes.

[Figure Five]

Discussion:

The broader purpose of this investigation was to demonstrate the utility of multiverse analysis techniques for evaluating the strength and generalisability of effects documented within neurological populations. For this purpose we employed a specific dataset recently used to investigate the relationship between acute visuo-spatial neglect and later functional/recovery outcomes[35]. With the multiverse analysis, we investigated the effect of choice of inclusion criteria, patient group definitions (here presence/absence of acute visuospatial neglect) and choice of outcome measure to represent ‘impact’. The multiverse analysis demonstrated that the predictive relationship between acute visuospatial neglect and poor recovery outcome is a particularly robust effect present across a range of outcome measures, patient groups, and scoring criteria. However, the variations in the strength of this effect across comparisons provides key information on the impact of researcher choices.

This analysis confirms that it is critically important to ensure neurological studies are adequately powered. In this investigation, comparisons employing sample sizes less than 100 were significantly less likely to detect effects which were reliably present when larger cohorts were considered. The risks associated with employing small sample sizes have been well-documented[36-38]. Conducting neurological investigations which are not powered to produce informative null results risks using valuable time and financial resources inefficiently.

Methodological choices relating to outcome measure used, stroke lateralisation, neglect categorisation criteria, and scoring method also created substantial variance within resultant effect sizes. Patients with neglect were found to have comparatively worse outcomes across each of the 16 outcome measures employed within this investigation. However, the average effect size of these comparisons varied dramatically between outcome measures. Neglect was found to have the weakest average impact on mood and the

communication subscale of the Stroke Impact Scale. In contrast, the strongest impact was found on measures assessing Activities of Daily Living. Interestingly, employing SIS ADL resulted in the strongest average effect sizes while employing the SIS IADL measure resulted in one of the weakest. The SIS ADL sub-measure aims to investigate mainly self-care activities and includes questions about personal hygiene, household chores, and continence. Conversely, the SIS IADL measure assesses more participation-based metrics and includes questions pertaining to social activities, hobbies, and interpersonal relationships[26]. The results of the present investigation suggests that neglect may impact more heavily on self-care activities than on relational or participation-based activities.

Comparisons which included all patients, regardless of lesion type resulted in stronger effect sizes than those which included patients with only right or left hemisphere damage. Similarly, the precise definition of neglect impairment employed within these comparisons was found to impact resultant effect sizes. Comparisons which exclusively considered egocentric neglect were found to result in weaker effect sizes than comparisons which also considered allocentric neglect. Most interestingly, analyses including only participants exhibiting both egocentric and allocentric neglect resulted in stronger effect sizes than those which included egocentric or allocentric patients alone. These findings align with previous research suggesting that the relationship between acute neglect and poor longer term outcome may not be uniform across all patients with neglect, but may instead vary significantly across different subtypes of neglect[9,24].

The multiverse results ultimately suggest that the strongest relationship between neglect and poor recovery outcome was found when neglect presence/absence was defined on the basis of demonstrating both ego-and allocentric neglect and activities of daily living were chosen as the outcome measure. It is possible this relates to an underlying factor of severity of neglect[10]. Additionally, in order to produce clinically useful information, it is critically

important to employ a representative sample of neglect including patients with right and left lesions and egocentric and allocentric impairment. The inclusion of more representative samples will facilitate results which can be generalised to the stroke population as a whole, rather than to only a specific subset of patients.

From this example, we extrapolate that multiverse analysis provides a novel method for evaluating the robustness of effects which could prove particularly useful for neurological studies as it enables researchers to distinguish between large-scale drivers of phenomena and delicate, ungeneralizable effects which are dependent on the exact methodological techniques employed[1], [8], [39]. Multiverse analysis enables researchers to isolate the impact of specific methodological decisions and evaluate the generalisability of their findings to wider patient populations.

Finally, with a growth of open data policies we suggest that multiverse analysis techniques have the potential to inform the design of future confirmatory studies, in an enhanced approach, where in addition to systematic reviews and meta analyses (which present the highest level of evidence[40]), multiverse analysis on existing datasets can present an enhanced opportunity to identify potential sources of bias within existing literature.

Conclusion:

This investigation demonstrates the utility of multiverse analysis techniques within neurological research with an example analysis of post-stroke impact of visuospatial neglect. The multiverse analysis makes explicit effects of researcher choices with respect to patient participant inclusion criteria, diagnostic criteria and primary outcome measure.

Competing Interests:

The authors report no competing interests.

Acknowledgements:

We would like to express our sincere gratitude and admiration to the late Prof Glyn W Humphreys, who initiated the OCS work and led the OCS-Care study. The OCS study was supported by the National Institute for Health Research Clinical Research Network. We also acknowledge the contributions to data collection and curation for the OCS data made by Ms Ellie Slavkova, Ms Grace Chiu, and Ms Romina Basting.

Funding:

This work was funded by Stroke Association UK awards to ND (TSA2015_LECT02; TSA 2011/02) and MJM (SA PGF 18\100031).

References:

- [1] S. Steegen, F. Tuerlinckx, A. Gelman, and W. Vanpaemel, "Increasing Transparency Through a Multiverse Analysis," *Perspect Psychol Sci*, vol. 11, no. 5, pp. 702–712, Sep. 2016, doi: 10.1177/1745691616658637.
- [2] M. F. Folstein, L. N. Robins, and J. E. Helzer, "The mini-mental state examination," *Archives of general psychiatry*, vol. 40, no. 7, pp. 812–812, 1983.
- [3] E. Mioshi, K. Dawson, J. Mitchell, R. Arnold, and J. R. Hodges, "The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening," *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, vol. 21, no. 11, pp. 1078–1085, 2006.
- [4] Z. S. Nasreddine *et al.*, "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.
- [5] B. L. Walter and J. L. Vitek, "Surgical treatment for Parkinson's disease," *The Lancet Neurology*, vol. 3, no. 12, pp. 719–728, Dec. 2004, doi: 10.1016/S1474-4422(04)00934-2.
- [6] J. Bernhardt *et al.*, "Agreed definitions and a shared vision for new standards in stroke recovery research: The Stroke Recovery and Rehabilitation Roundtable taskforce," *Int J Stroke*, vol. 12, no. 5, pp. 444–450, Jul. 2017, doi: 10.1177/1747493017711816.
- [7] Y. Yamada, "How to Crack Pre-registration: Toward Transparent and Open Science," *Front. Psychol.*, vol. 9, 2018, doi: 10.3389/fpsyg.2018.01831.
- [8] P. Dragicevic, Y. Jansen, A. Sarma, M. Kay, and F. Chevalier, "Increasing the transparency of research papers with explorable multiverse analyses," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–15.

- [9] M. J. Moore, K. Vancleef, J. M. Riddoch, C. R. Gillebert, and N. Demeyere, "Visuospatial neglect: Recovery & Functional outcome after 6 months.," *NEUROREHABILITATION AND NEURAL REPAIR*, Under Review.
- [10] N. Demeyere and C. R. Gillebert, "Ego-and allocentric visuospatial neglect: Dissociations, prevalence, and laterality in acute stroke.," *Neuropsychology*, vol. 33, no. 4, p. 490, 2019.
- [11] A. Parton, P. Malhotra, and M. Husain, "Hemispatial neglect," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 1, pp. 13–21, Jan. 2004.
- [12] W. L. Bickerton, D. Samson, J. Williamson, and G. W. Humphreys, "Separating forms of neglect using the Apples Test: Validation and functional prediction in chronic and acute stroke.," *Neuropsychology*, vol. 25, no. 5, pp. 567–580, 2011, doi: 10.1037/a0023501.
- [13] L. R. Cherney, A. S. Halper, C. M. Kwasnica, R. L. Harvey, and M. Zhang, "Recovery of functional status after right hemisphere stroke: Relationship with unilateral neglect," *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 3, pp. 322–328, Mar. 2001, doi: 10.1053/apmr.2001.21511.
- [14] M. Jehkonen *et al.*, "Visual neglect as a predictor of functional outcome one year after stroke," *Acta Neurologica Scandinavica*, vol. 101, no. 3, pp. 195–201, Mar. 2000, doi: 10.1034/j.1600-0404.2000.101003195.x.
- [15] S. Stone, P. Patel, R. Greenwood, and P. Halligan, "Measuring visual neglect in acute stroke and predicting its recovery: the visual neglect recovery index.," *J Neurol Neurosurg Psychiatry*, vol. 55, no. 6, pp. 431–436, Jun. 1992, doi: 10.1136/jnnp.55.6.431.
- [16] A. F. T. Ten Brink, J. H. Verwer, J. M. Biesbroek, J. M. A. Visser-Meily, and T. C. W. Nijboer, "Differences between left- and right-sided neglect revisited: A large cohort study across multiple domains," *Journal of Clinical and Experimental Neuropsychology*, vol. 39, no. 7, pp. 707–723, Aug. 2017, doi: 10.1080/13803395.2016.1262333.
- [17] G. M. S. Nys *et al.*, "Early cognitive impairment predicts long-term depressive symptoms and quality of life after stroke," *Journal of the Neurological Sciences*, vol. 247, no. 2, pp. 149–156, Sep. 2006, doi: 10.1016/j.jns.2006.04.005.
- [18] L. J. Buxbaum *et al.*, "Hemispatial neglect: Subtypes, neuroanatomy, and disability," *Neurology*, vol. 62, no. 5, pp. 749–756, Mar. 2004, doi: 10.1212/01.WNL.0000113730.73031.F4.
- [19] A. E. Hillis and A. Caramazza, "A framework for interpreting distinct patterns of hemispatial neglect," *Neurocase*, vol. 1, no. 3, pp. 189–207, Jul. 1995, doi: 10.1080/13554799508402364.
- [20] M. Chechlacz, P. Rotshtein, and G. W. Humphreys, "Neuroanatomical Dissections of Unilateral Visual Neglect Symptoms: ALE Meta-Analysis of Lesion-Symptom Mapping," *Front Hum Neurosci*, vol. 6, Aug. 2012, doi: 10.3389/fnhum.2012.00230.
- [21] R. Ptak, M. Di Pietro, and A. Schnider, "The neural correlates of object-centered processing in reading: A lesion study of neglect dyslexia," *Neuropsychologia*, vol. 50, no. 6, pp. 1142–1150, May 2012, doi: 10.1016/j.neuropsychologia.2011.09.036.
- [22] H. Ota, T. Fujii, K. Suzuki, R. Fukatsu, and A. Yamadori, "Dissociation of body-centered and stimulus-centered representations in unilateral neglect," *Neurology*, vol. 57, no. 11, pp. 2064–2069, Nov. 2001, doi: 10.1212/WNL.57.11.2064.
- [23] N. Demeyere, M. J. Riddoch, E. D. Slavkova, W.-L. Bickerton, and G. W. Humphreys, "The Oxford Cognitive Screen (OCS): Validation of a stroke-specific short cognitive

- screening tool.,” *Psychological Assessment*, vol. 27, no. 3, pp. 883–894, 2015, doi: 10.1037/pas0000082.
- [24] W.-L. Bickerton, D. Samson, J. Williamson, and G. W. Humphreys, *Separating Forms of Neglect Using the Apples Test: Validation and Functional Prediction in Chronic and Acute Stroke*, vol. 25. 2011.
- [25] N. Demeyere and C. Gillebert, “Dissociating ego- and allocentric neglect after stroke: prevalence, laterality and outcome predictors,” *bioRxiv*, p. 200386, Oct. 2017, doi: 10.1101/200386.
- [26] Duncan Pamela W., Wallace Dennis, Lai Sue Min, Johnson Dallas, Embretson Susan, and Laster Louise Jacobs, “The Stroke Impact Scale Version 2.0,” *Stroke*, vol. 30, no. 10, pp. 2131–2140, Oct. 1999, doi: 10.1161/01.STR.30.10.2131.
- [27] P. Lyden, “Using the National Institutes of Health Stroke Scale: A Cautionary Tale,” *Stroke*, vol. 48, no. 2, pp. 513–519, Feb. 2017, doi: 10.1161/STROKEAHA.116.015434.
- [28] H. Al-Janabi, T. N. Flynn, and J. Coast, “Development of a self-report measure of capability wellbeing for adults: the ICECAP-A,” *Quality of Life Research*, vol. 21, no. 1, pp. 167–176, 2012.
- [29] I. Bjelland, A. A. Dahl, T. T. Haug, and D. Neckelmann, “The validity of the Hospital Anxiety and Depression Scale: an updated literature review,” *Journal of psychosomatic research*, vol. 52, no. 2, pp. 69–77, 2002.
- [30] B. C. Robinson, “Validation of a caregiver strain index,” *Journal of gerontology*, vol. 38, no. 3, pp. 344–348, 1983.
- [31] T. J. Quinn, P. Langhorne, and D. J. Stott, “Barthel Index for Stroke Trials: Development, Properties, and Application,” *Stroke*, vol. 42, no. 4, pp. 1146–1151, Apr. 2011, doi: 10.1161/STROKEAHA.110.598540.
- [32] U. Simonsohn, J. P. Simmons, and L. D. Nelson, “Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2694998, Oct. 2019. doi: 10.2139/ssrn.2694998.
- [33] A. Kühberger, A. Fritz, and T. Scherndl, “Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size,” *PLoS ONE*, vol. 9, no. 9, p. e105825, Sep. 2014, doi: 10.1371/journal.pone.0105825.
- [34] “Sailing From the Seas of Chaos Into the Corridor of Stability: Practical Recommendations to Increase the Informational Value of Studies - Daniël Lakens, Ellen R. K. Evers, 2014.” <https://journals.sagepub.com/doi/full/10.1177/1745691614528520> (accessed Mar. 10, 2020).
- [35] M. J. Moore, K. Vancleef, M. J. Riddoch, C. R. Gillebert, and N. Demeyere, “Visuospatial Neglect: Functional Outcome and Recovery at 6 Months,” *Journal of Clinical and Experimental Neuropsychology*, Under Review.
- [36] J. Cohen, “A power primer.,” *Psychological bulletin*, vol. 112, no. 1, p. 155, 1992.
- [37] S. E. Maxwell, “The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies,” *Psychological Methods*, vol. 9, no. 2, pp. 147–163, 2004, doi: 10.1037/1082-989X.9.2.147.
- [38] R. M. Turner, S. M. Bird, and J. P. T. Higgins, “The Impact of Study Size on Meta-analyses: Examination of Underpowered Studies in Cochrane Reviews,” *PLOS ONE*, vol. 8, no. 3, p. e59202, Mar. 2013, doi: 10.1371/journal.pone.0059202.
- [39] A. Orben and A. K. Przybylski, “Teenage sleep and technology engagement across the week,” *PeerJ*, vol. 8, p. e8427, Jan. 2020, doi: 10.7717/peerj.8427.
- [40] P. B. Burns, R. J. Rohrich, and K. C. Chung, “The Levels of Evidence and their role in Evidence-Based Medicine,” *Plast Reconstr Surg*, vol. 128, no. 1, pp. 305–310, Jul. 2011, doi: 10.1097/PRS.0b013e318219c171.

Figure Captions:

Figure 1: The OCS cancellation task. Panel A presents examples of significant egocentric and allocentric neglect impairment. Panel B presents a visualisation of how the Centre of Cancellation asymmetry metric is calculated.

Figure 2: The upper panel presents a specification curve analysis presenting the effect sizes of all experimental and null comparisons ranked in order of size on the x-axis. Significant comparisons are represented in blue, insignificant results are visualised in red, and null curve comparisons are represented in grey. Dotted lines denote the 25th and 75th quantile of the null curve. 99.9% of experimental comparisons fell below the 25th quantile of the null curve, suggesting that neglect illustrates that the relationship between neglect and poor outcome is highly robust and significant across many possible methodological combinations. The lower panels visualise each specific combination of methodological comparisons used in each individual comparison. The average effect size and standard deviation of comparisons using each methodological option are listed to the right of each row. CS = Carer Strain, QOL = Quality of Life, ADL 2 = SIS Instrumental ADL, Com. = SIS communication, AffHand = SIS Affected Hand, Weak = SIS Physical Weakness, Not BL = Not Bilateral, Ego = Egocentric, Allo = Allocentric.

Figure 3: The effect sizes of each conducted comparison grouped by sample size (x-axis). Significant comparisons are presented in blue while insignificant comparisons are in red. Black dots represent the mean effect size of each group. The number of comparisons and overall percent significant results within each group are presented.

Figure 4: The effect sizes of each conducted comparison grouped by outcome measure (x-axis). Significant comparisons are presented in blue while insignificant comparisons are in red. Black dots represent the mean effect size of each group. The number of comparisons and overall percent significant results within each group are presented. Notably, all outcome measures resulted in an average negative effect size, indicating a robust relationship between the occurrence of acute neglect and poor outcome.

Figure 5: A visualisation of the impact of individual methodological choices on effect size. Each box represents the combination of two specific methodological choices listed on the x and y axes. Box colour represents the average effect size of comparisons which employed each specific combination of choices (See Key). Boxes are labelled with average sample size. The average effect size of each choice across all conducted comparisons is also reported. Red total values represent average effect sizes which are significantly higher than the overall mean effect and blue values represent average effects which are significantly lower than the mean effect as determined by t-tests. Only comparisons which are sufficiently powered to detect the average significant effect size (-0.56) are included in this figure ($n > 105$).

A

Egocentric Neglect

Allocentric Neglect

B

Centre of Cancellation = $\frac{\sum \text{Weights}}{N \text{ Targets Found}}$

Effect Size

medRxiv preprint doi: <https://doi.org/10.1101/2021.03.29.21254222>; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

● = significant ● = not significant

mean n	146	203	209	208	207	208	209	204	206	209	208	217	223	235	237	69
% signif.	30.3	58.1	40.0	19.9	24.7	71.0	41.7	29.7	21.0	62.7	59.8	1.77	6.31	43.5	35.3	4.95
	NIHSS	PhysWeak (SIS)	Memory (SIS)	Mood (SIS)	Com. (SIS)	ADL (SIS)	Mobility (SIS)	AffHand (SIS)	ADL 2 (SIS)	Total (SIS)	ADL Total (SIS)	Anxiety (HADS)	Depression (HADS)	Barthel	QOL	CS

Outcome Measure

Effect Size

Outcome Measure	Total	-0.52 (36)
	Weak	-0.43 (31)
	Mood	-0.25 (20)
	Mobility	-0.37 (20)
	Memory	-0.36 (30)
	Com.	-0.27 (41)
	Aff. Hand	-0.33 (27)
	ADL Total	-0.44 (32)
	ADL 2	-0.19 (21)
	ADL	-0.57 (40)
	QOL	-0.31 (16)
	NIHSS	-0.42 (45)
	Depress.	-0.09 (25)
	CS	-0.22 (35)
	Barthel	-0.39 (24)
	Anxiety	-0.11 (20)

Lesion Side	Visible	-0.33 (24)
	Right	-0.30 (30)
	Left	-0.35 (53)
	Bilateral	-0.32 (26)
	Any	-0.35 (23)

medRxiv preprint doi: <https://doi.org/10.1101/2021.03.29.21254222>; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Stroke Type	NOTA	-0.31 (21)
	Known	-0.31 (21)
	Ischem.	-0.29 (22)
	Haem.	-0.46 (60)
	Any	-0.29 (21)

Assessment Time	Any	-0.32 (31)
	Known	-0.34 (31)
	< 62d	-0.34 (15)
	< 31d	-0.34 (33)
	< 15d	-0.32 (35)
	< 8d	-0.30 (36)

Score Neg. Type	Ego + Allo	-0.38 (35)
	Any	-0.33 (28)
	Allo Only	-0.30 (24)
	Ego Only	-0.30 (41)

Score	No Min.	-0.30 (32)
	Min.	-0.35 (33)