The Clinical Neuropsychologist

ISSN: 1385-4046 (Print) 1744-4144 (Online) Journal homepage: http://www.tandfonline.com/loi/ntcn20

Measuring and diagnosing unilateral neglect: a standardized statistical procedure

Alessio Toraldo, Cristian Romaniello & Paolo Sommaruga

THE CLINICAL NEUROPSYCHOLOGIST, 2017
https://doi.org/10.1080/13854046.2017.1349181

Routledge
Taylor & Francis Group

Check for updates

Measuring and diagnosing unilateral neglect: a standardized statistical procedure

Alessio Toraldo[a], Cristian Romaniello[a] and Paolo Sommaruga[b]

[a]Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy; [b]Technology Operative Center, Reti Televisive Italiane, Segrate, Italy

ABSTRACT

Objective: Unilateral neglect is usually investigated by adminstering stimuli (targets) in different positions, with targets being responded to by the patient (Hit) or omitted. In spite of this homogeneity of data type, neglect indices and diagnostic criteria vary considerably, causing inconsistencies in both clinical and experimental settings. We aimed at deriving a standard analysis which would apply to all tasks sharing this data form. **Methods:** A-priori theoretical reasoning demonstrated that the mean position of Hits in space (MPH) is an optimal index for correctly diagnosing and quantifying neglect. Crucially MPH eliminates the confounding effects of deficits that are different from neglect (non-lateral) but which decrease Hit rate. We ran a Monte Carlo study to assess MPH's (so far overlooked) statistical behavior as a function of numbers of targets and Hits. **Results:** While average MPH was indeed insensitive to non-lateral deficits, MPH's variance (like that of all other neglect indices) increased dramatically with increasing non-lateral deficits. This instability would lead to alarmingly high false-positive rates (FPRs) when applying a classical diagnostic procedure that compares one patient with a control sample. We solved the problem by developing an equation that takes into account MPH instability and provides correct cut-offs and close-to-nominal FPRs, even without control subjects. We developed a computerized program which, given the raw data, yields the MPH, a *z*-score and a *p*-value. **Conclusions:** We provided a standard method that allows clinical and experimental neuropsychologists to diagnose and measure neglect in a consistent way across the vast majority of tasks.

Abbreviations: MPH, Mean Position of Hits; MPO, Mean Position of Omissions; MPT, Mean Position of Targets; MOH, Mean Ordinal position of Hits; MdnPH, Median Position of Hits; FPR, false positive rate (α probability of type-I error); FNR, false negative rate (β probability of type-II error); LCR-adjusted, Left-Center-Right-adjusted; C-adjusted, Center-adjusted (or centered); CoC, Center of Cancellation; SD, expected standard deviation of MPH (unless otherwise stated); BF, Bayes Factor; *H*, number of Hits; *T*, number of targets; *G*, number of target clusters; FA, False Alarm (on catch trials); CR, Correct Rejections (on catch trials).

ARTICLE HISTORY
Received 12 January 2017
Accepted 26 June 2017

KEYWORDS
Unilateral neglect; diagnosis; cancellation; visual search; Mean Position of Hits

CONTACT Alessio Toraldo alessio.toraldo@unipv.it, alessio.toraldo@gmail.com, alessio@toraldo.it

The underlying research materials for this article can be accessed at <http://psicologia.unipv.it/toraldo/mean-position-of-hits.htm>

Introduction

Unilateral neglect has been widely investigated both for theoretical and clinical reasons across many decades. This enterprise has involved a huge variety of clinical and experimental tests. The most common paper-and-pencil test is cancellation, which is administered in several different forms (Albert, 1973; Diller & Weinberg, 1977; Edgworth, Robertson, & MacMillan, 1998; Gauthier, Dehaut, & Joanette, 1989; Friedman, 1992, etc.). On these tests, the patient is asked to search for specific visual stimuli (targets), that are homogeneously distributed across (typically) a landscape-oriented A4 or A3 paper sheet, and cross them all out with a pencil. Targets may be presented in isolation (Albert), or intermixed with distractors (Diller & Weinberg); they can be lines (Albert), letters (Diller & Weinberg) or silhouettes of objects, like stars (Friedman), and be presented in separate rows (Diller & Weinberg) or in a homogeneous, unstructured field (Friedman). Beyond the visual modality, 'haptic' versions of the cancellation test have been developed in which the subject has to remove a series of objects, while blindfolded (e.g. Bisiach, Capitani, & Porta, 1985). In the tactile modality, separate stimuli are presented on the patient's skin (e.g. Pitteri, Venneri, Meneghello, & Priftis, 2013). In the auditory modality neglect patients have been typically tested by presenting single auditory stimuli, one at a time in different positions (e.g. De Renzi, Gentilini, & Barbieri, 1989). There are also tasks of single-target detection in the visual modality (however, if fixation is kept at a stable point, like in Posner, Walker, Friedrich, and Rafal's (1984) paradigm or in visual field examinations like static perimetry, the test is also sensitive to visual scotomas, and not only to neglect). Tasks involving recall from short-term memory have also been proposed (e.g. Moreh, Malkinson, Zohary, & Soroker, 2014).

This large variety of behavioral tests is an objective problem especially in clinical settings. Clinical neuropsychologists often face the challenge of how to compare performance obtained from different tests or analyzed with different diagnostic criteria. Procedures typically vary even between hospitals of the same region, making the evolution of a patient's deficit difficult to assess because of, for instance, the discrepancies between the tools used in the neurological ward where the patient was admitted in the acute phase and those used in the rehabilitation unit.

While pursuing uniformity in the choice of the behavioral tests across hospitals worldwide is an impossible enterprise, we believe that uniformity might be achieved at least on the choice of the mathematical index and of the relevant statistical procedures. These aspects, which have been largely overlooked so far, are the purpose of the present work.

Indeed, all of the neglect tasks listed above and many more in a huge literature share one relevant detail, namely the exact form of the raw data obtained from a patient. The behavioral output is a set of positions homogeneously distributed across one spatial dimension (typically the horizontal one), each of which is associated with a target that was either successfully processed[1] (Hit, 1) or not (Omission, 0) – briefly, the 'Hit/Miss-by-position' data form. In spite of the data form being exactly the same, *different* indices and procedures have been used for diagnosing and quantifying neglect in different tasks (and often also for the same task). We will propose a way out of this anomaly by deriving a single, theoretically-founded neglect index that best captures the essence of a patient's performance. This would be beneficial to both the scientific and the clinical community working on neglect because it would provide them with a standard way to analyze most data. Hence, no matter the tested sensory modality (visual, auditory, tactile, etc.), the type of response signaling successful target processing (verbal, reaching movement, eye movement, dichotomized response times, dichotomized physiological responses like ERP, fMRI, skin conductance, etc.), the role of extra cognitive operations (presence/absence of distractors, etc.) and even the type of task (active search within a homogeneous field of simultaneously presented stimuli, like in visual or haptic cancellation, or explicit/implicit detection of single stimuli presented one at a time in different locations, like in Posner's paradigm or static perimetry, or short/long-term memory recall of previously inspected stimulus arrays, etc.), a single, standard measure and diagnostic procedure would be available.

What this paper is *not* about is any neglect task producing a raw data form that is different from the Hit/Miss-by-position one. For example, in line bisection or straight-ahead pointing (Karnath, 1997), a single (invisible) target is pointed to repeatedly, and scores vary along a continuous measure of distance, not within a Hit/Miss dichotomy. Eye movement recording during free inspection of a visual scene (Fruhmann-Berger & Karnath, 2005) is not a good candidate either, because there is no set of pre-determined targets that the subject is required to process. Identical remarks hold for many famous long-term memory recall tasks in which the patient is asked to mentally visualize a known place and describe it (Bisiach & Luzzatti, 1978). The Landmark task (e.g. Toraldo, Laiacona, Pagani, Mandelli, & Capitani, 2014; Toraldo, McIntosh, Dijkerman, & Milner, 2004) does have a stimulus position varying across trials, but each trial produces a judgment of relative position of the stimulus, and not a Hit/Miss score. Tests for neglect dyslexia using text or single (non)word reading are a special case. Notwithstanding the fact that there is no explicit detection of separate stimuli, single letters along a word can be considered as targets varying in their horizontal position whose successful processing can be inferred from the verbal response. However, the effective processing of each letter is masked by subsequent lexical/sub-lexical elaboration, for instance, by the opacity of the grapheme-to-phoneme conversion rules. Hence neglect dyslexia tests will be the subject of future, dedicated work. One last remark is that in many tasks, stimuli/trials that are to be responded to ('targets') are intermixed with stimuli/trials that are *not* to be responded to ('distractors', or 'catch trials' etc.). Thus, one has 'False Alarms' (FA) and 'Correct Rejections' (CR) along with Hits and Misses. This paper only treats Hits and Misses – future work will address the issue of how to include FA and CR in the analysis.

The structure of the paper is as follows. (1) We show the inadequacy of the most widespread indices for measuring neglect on Hit/Miss-by-position tasks (section 'Classical mathematical indices of neglect'). (2) We then propose theoretically-founded *a priori* reasoning that justifies the choice of an optimal index, which is the Mean Position of Hits across physical space, MPH ('Deriving a better neglect index'). (3) We discuss the statistical properties of MPH and their effects on diagnostic procedures, which we could simulate in a Monte Carlo study ('Diagnosing neglect with the MPH index'). (4) We provide downloadable software for MPH computation and for obtaining statistically correct diagnosis of neglect, which is valid whatever task was used to generate the data ('Software for automatic computation'). Step (3) is critical insofar as it treats a serious problem with neglect diagnosis which, to our knowledge, has been generally overlooked: all neglect indices (MPH included) get increasingly unstable as Hit rate decreases, causing serious inflation of the probability of false positives. We solved the problem by including MPH instability in the statistical model used for diagnosis (3 and 4).

The theoretical work detailed in (1), (2) and in part of (3) was carried out by one of us (AT) across several years (2000–2007). This work led to his lab's regular use of MPH (which was initially called 'Centre of Gravity') in both clinical and experimental activities (Bottini, Gandola, Scarpa, Toraldo, & Zanardi, 2005; Gandola et al., 2007, 2013). Many of the technical-statistical-mathematical details were omitted from this paper for the sake of brevity. The interested reader can access the 'Website Material' (psicologia.unipv.it/toraldo/mean-position-of-hits.htm) or contact the authors. Readers who are only concerned with the practical application of the method can skip sections '"Classical" mathematical indices of neglect', 'Deriving a better index', 'Diagnosing neglect with the MPH index' and directly read section 'Software for automatic computation'.

'Classical' mathematical indices of neglect

Two indices are definitely the most widespread for Hit/Miss-by-position data: the difference between the counts (or percentages) of Hits on the two halves of the display (*R-L*) and the *Overall Hit Rate* [*Accuracy*, independently of position: Hits/(Hits + Misses)]. Another less frequent index is the *Proportion of Hits on one half*, for instance, Hits on the Left divided by overall Hit count (L/TOT, Halligan et al., 1991). Most typically neglect is diagnosed if the indices surpass some (often arbitrary) threshold, like a R-L difference of 2 or 3 counts. Irrespective of cut-offs, such indices all fail to distinguish cases in which neglect severity is obviously different. Table 1 allows for a set of systematic comparisons between performances which illustrate the point. Patient 1 (P1) detected only the targets of the right-most quarter of the display, P2 only the right half, P3 omitted only the left-most quarter, and P4 omitted half the targets but homogeneously so across all of the display. Everybody would agree that P1, P2, P3, and P4 are in decreasing order of neglect severity. Yet, none of the three cited indices correctly reflects this order. R-L gives the same score (+50%) to both P1 and P3, L/TOT matches patients P1 and P2 (score = 0), and *Accuracy* matches P2 and P4 (score = 50%). These misclassifications are gross: for example, the failure by *Accuracy* to record the difference between a patient who detects all (and only) the targets in the right half (P2) and a patient who detects half the targets in all sectors of space, without lateral bias (P4) is striking. A better index is needed.[2]

One possibility would be that of scrutinizing an immense literature in search of mathematical objects that correctly classify patients P1–P4. However, P1–P4 are just a small subset of all possible performances and represent rather extreme cases: a correct classifier of the

Table 1. Four imaginary patients (P1–P4) whose performance is graphically represented on the left side of the table are compared.

		Hit rate				Hit rate (%)		Neglect scores			
		Display quarters				Halves					
		1	2	3	4	L	R	R-L difference (%)	L/TOT ratio	Overall Hit rate (%)	MPH, MdnPH, Mid-Range
Patient	P1					0	50	+50	0	25	4
	P2					0	100	+100	0	50	3.5
	P3					50	100	+50	.333	75	3
	P4					50	50	0	.5	50	2.5

Notes: The four quarters of the display (1–4, left to right) where the patient searched for targets are shown, and their Hit rate is represented in shades of gray, from white (no target detected), to black (all of the targets detected). The three neglect indices discussed in the text are computed; patients P1–P4 are assumed to be ordered for decreasing neglect severity, hence failures to correctly differentiate pairs of patients are shown by underscoring their values. MPH, MdnPH, Mid-Range are central tendency measures discussed later in the paper.

P1–P4 order might well fail on new hypothetical pairs. Hence, we reasoned that a more general analysis of all possible (or plausible) performances would be decisive at this stage. We have tried to do so in the next section.

Deriving a better index

Looking at performance curves: what features measure neglect?

In this section we aimed to find a correspondence between a complete description of the performance of a patient and the putative degree of neglect affecting it – what we wish to measure. The raw performance of a patient is best formalized as the function relating the horizontal position of targets to Hit rate, which is usually modeled as a logistic curve. In Figure 1, several examples of this function are shown, with target position on the horizontal axis (ranging conventionally from −.5, left-most target, to .5, right-most target) and Hit rate on the vertical (ranging 0–1). The examples are grouped in such a way that, within each plot, only one parameter of the curve varies. The varying parameter assumes three different values

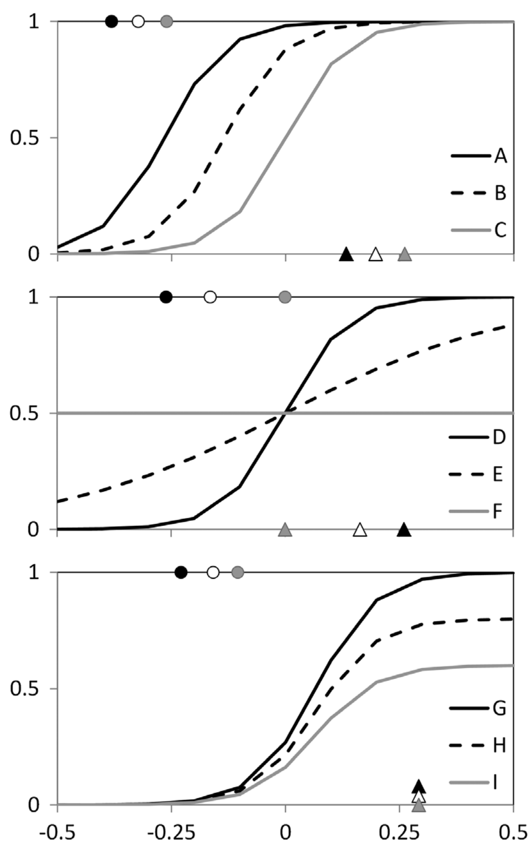

Figure 1. Physical position ranging from −.5 (left-most target) to .5 (right-most target) is plotted against Hit rate (0–1) for nine imaginary patients.
Notes: Triangles along the plots' horizontal axes show the Mean Position of Hits (MPH): black, relative to the black solid curve; white, relative to the dashed curve; gray, relative to the gray curve. MPHs are identical in patients G–I, so the triangles overlap. Circles along the plots' upper borders show the Mean Position of Omissions (MPO) – color conventions as before.

in each plot, creating the performance curves of three imaginary patients. Hence, if the three patients in a plot seem[3] to show the same degree of neglect, that parameter will be irrelevant for a good neglect measure; else, if the three patients seem to show different degrees of neglect, that parameter will be relevant for a good neglect measure.

Patients A, B, and C all show a left-ward drop of Hit rate – the *slope* of the drop is the same in all the three cases, but the drop occurs at different *locations* in space. In patients D, E, and F, the *location* of the drop is the same, but the *slope* is different. Finally, patients G, H, and I show a drop in the same *location* in space, but each starts from a different maximum performance (*ceiling*). Accordingly, we find the three parameters of the curve: (i) the *ceiling* (the maximum Hit rate achieved anywhere by the patient), (ii) how steep the drop is, starting from the ceiling (the *slope*), and (iii) the *location* of the drop (where in space the patient's Hit rate is exactly half the ceiling).[4]

We are now ready to see what differences in neglect severity follow variations in each parameter.

None would doubt that patients A, B, and C have different neglect levels: A has the mildest, C the most severe form of neglect. Therefore, *location* is a relevant parameter.

The change from patient D to E to F is a progressive reduction in the *slope*, with *location* of the drop being constant.[5] All would agree that patient F does not have neglect, because there is no left–right imbalance, so patient F must be the 'zero-neglect' pole of the progression. The deduction is that F has less severe neglect than E, and E has less severe neglect than D. Therefore, also *slope* is a relevant parameter: the steeper the curve, the more severe the neglect.

The really critical point is to clarify that patients G, H, and I should be assumed to have the same degree of neglect. Curves such as those in G, H, and I would be obtained if, for example, a *single* patient with neglect were tested in three separate days, having three different levels of some deficit *other* than neglect ('non-lateral') that decreases his/her overall performance, while the neglect remained constant. Hence, plots of that kind should all be assumed to correspond to the same level of neglect severity. There are several examples of non-lateral deficits that can co-occur with neglect because of the lesion, and which would affect performance: for example, on a cancellation task, amblyopia, visual agnosia, optic ataxia (if free vision is allowed), dyslexia (if targets are alphanumeric characters), deficits of executive functions (the patient cannot focus on the task and is 'captured' by irrelevant stimuli like the distractors), lack of concentration or of motivation, deficit of vigilance, would all decrease the probability of target detection. The exact nature of additional deficits is irrelevant, their only critical feature is that they affect the processing of targets *irrespective of their position along the studied dimension*.[6]

The way in which neglect and a non-lateral deficit combine is illustrated by cases G, H, and I. Patient G represents a case of 'pure' neglect: performance is perfect on the far right, so there cannot be any contribution by non-lateral deficits. If one 'adds' some non-lateral deficits, passing from patient G to H or I, the curve will be linearly compressed towards the floor. This happens because we need to multiply the probabilities of successful processing by operations that produce lateral deficits (neglect) when lesioned and operations that cause non-lateral deficits when lesioned.[7] Wrapping up, all patients whose Position vs. Hit rate plot is identical under simple multiplicative transformations should be assumed to have identical neglect severity. This is a mandatory assumption, implicit in the use of all Hit/Miss-by-position tasks (i.e. in the vast majority of neglect tests). If this assumption were false, it would be impossible to distinguish lateral from non-lateral deficits.[8]

Since patients G, H, and I differ only for the *ceiling* parameter, and we concluded that they should be assumed to have identical degree of neglect, the *ceiling* parameter must be assumed to be irrelevant for any measure of neglect severity.

To summarize, while *ceiling* must be left out of any neglect measure, *slope* and *location* parameters are relevant. Could just one of them be used directly as a good index? Certainly not, because either parameter would fail to distinguish cases that differ for the other parameter. Thus, *slope* would fail to tell the different neglect severities of patients A–C (Butler et al., 2004, used this parameter), and *location* would fail to tell the different neglect degrees of patients D–F. So, one must combine information from both parameters.

Looking at distributions of Hits across space suggests some valid measures

We found a straightforward way to combine information from both parameters. The trick is to pay attention not to the curves, but to the areas under the curves in Figure 1. These areas are the statistical distributions of the Hits across the space of the display – the histograms of Hits across positions. Distributions can be described by a small number of features, with the main ones being central tendency (where the Hits are in the display) and dispersion (how variable the positions of Hits are in the display). Guided by this change in perspective, we realized that *any* measure of central tendency as well as *any* measure of dispersion would be sensitive to both variations in *slope* and variations in *location*, in exactly the way we wished, that is, reflecting neglect severity order. By looking at plots A–C in Figure 1 – the set of patients varying only for *location* – it is clear that both the mean and the variance of the distribution change between patients. Mean and variance also change between patients D–F, those varying only for *slope*, and critically, mean and variance do *not* vary between cases G–I, those in whom both *slope* and *location* are fixed (neglect severity is constant). The distributions of patients G–I look like a hill that got progressively 'compressed' downwards (from G to H to I): if you make a hill shallower, its mean and variance stay the same.[9]

In a thorough Monte Carlo study (detailed in the Website Material), we explored the statistical characteristics of a number of central tendency and dispersion measures. Briefly, dispersion measures were excluded because they are also sensitive to 'double neglect' (neglect on both sides) which would be often mistaken for severe left or right neglect, and because they distribute in a non-Gaussian way. As for central tendency measures, we studied the Mean Position of Hits (MPH), the Median Position of Hits (MdnPH), and the *Mid-Range* (the mean between the positions of the left-most and right-most Hits). The third index proved suboptimal because its distributions are often markedly leptokurtic and because it is disastrously insensitive to mild neglect. The distributions of mean (MPH) and median (MdnPH) are nicely close to Gaussian (the mean is slightly better, Kurtosis = −.28 vs. −.48), however, the mean is definitely more efficient: its standard deviation (SD) is about 1.53 times smaller than that of the median.

Our search came to an end: the Mean Position of Hits, MPH, proved to be the best available index from both the theoretical and the statistical viewpoint.

Examples of use of MPH

MPH has already been used with neglect patients. For instance, in cancellation tests, Binder, Marshall, Lazar, Benjamin, and Mohr (1992) computed the arithmetic mean of the position of cancelled targets across the display, an example of MPH. Relying on the above theoretical analysis, one of us (AT) with many colleagues have been using MPH in cancellation and visual search tasks for several years in clinical settings (Niguarda Hospital, Milan, since 2004; Maugeri Hospital, Pavia, 2005–2012) and in experimental work on neglect-perseveration (Gandola et al., 2007, 2013). More recently (since 2010) those labs have analyzed MPH with the statistical procedure explained below. Rorden and Karnath (2010) proposed a large-scale use of the 'Centre of Cancellation' (CoC), the MPH in a cancellation task, and developed a user-friendly program that computes it from a scan of the to-be-explored display. What is still missing in the published literature is a correct statistical test for diagnosing neglect with the MPH. This is the topic of the next-but-one 'Diagnosing neglect with the MPH index' section.

What if one considers Omissions instead of Hits?

The reader might wonder whether the Mean Position of Hits (MPH) is related to the Mean Position of Omissions (MPO). If both of them are standardized and 'C-adjusted' (see the 'Software for automatic computation' section) there is an exact relationship between them: MPH = –MPO[$(T - H)/H$] where T = number of targets and H = number of Hits. However, critically, MPH and MPO are *not* equivalent in terms of their ability to distinguish different degrees of neglect. In the crucial G–I patients (Figure 1), which we assumed to have the same degree of neglect, MPHs are identical, while MPOs are different. This can be understood by turning Figure 1 upside-down: the areas under the curves will then show the Omission rate distributions (MPOs are shown as circles). Hence, MPH is to be preferred to MPO as a measure of neglect severity. Despite these discrepancies, both (C-adjusted) MPH and MPO lead to identical diagnostic inferences (identical p-values, see Website Material).

Diagnosing neglect with the MPH index

In this section, we summarize the work we carried out to derive a correct statistical procedure for diagnosing neglect with MPH. During this process we confirmed that MPH suffers, as any other neglect index, from a problem that has been surprisingly overlooked in the literature so far, and which can produce large proportions of false positives. The 'neglected' problem is that identical values of an index do not have the same diagnostic meaning if they are associated with different overall Hit rates. Take the widespread R-L difference in Hit counts. It is common practice to diagnose neglect on the Albert (1973) task if a difference of at least 2 Hits is recorded between the halves. However, a difference of 2 is much more likely to be just the effect of a random fluctuation (a false positive) when overall Hit rate is close to 50% (e.g. 8/10 vs. 10/18) than when it is far from it (e.g. 16/18 vs. 18/18). We performed a large Monte Carlo study to show that a similar[10] problem also affects MPH.

Monte Carlo simulations

In each simulation we generated 10,000 random distributions of H Hits in T targets equispaced along a reference horizontal space. In different simulations T could be 10, 20, 50, 100, or 150 and H could be 1, 2, 3, $T - 1$, $T - 2$, $T - 3$ or fractions of T (from $.1T$ to $.9T$ in steps of $.1T$). This corresponds to generating data from virtual subjects without neglect (isoprobability

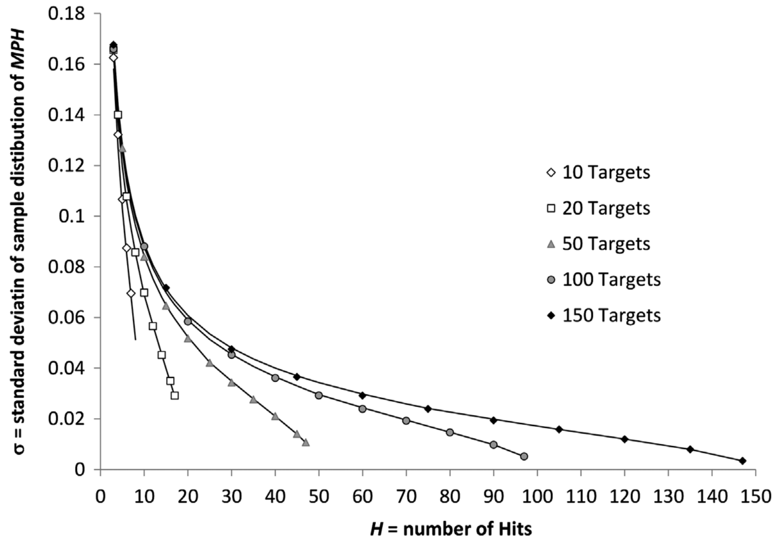

Figure 2. Standard deviation of the sample MPH distribution as a function of Hit count (horizontal axis) and number of targets (different plots).
Notes: Dots show values obtained from Monte Carlo simulations of 10,000 samples each, generated under null hypothesis of no lateral bias (Hit isoprobability across all positions in the display, with *T* targets in *T* equispaced positions); curves show the fit obtained with the Equation reported in the Appendix.

of detection across space) and with various degrees of non-lateral deficits (i.e. of overall Hit rate). Additional simulations were run to study the effect of having targets grouped in fewer positions, according to all possible pairs of factors of *T* (e.g. for 10 targets we additionally simulated the cases of 5 positions, each repeated twice, and 2 positions, each repeated 5 times). Hence, each simulation provided 10,000 MPHs whose distributions could be analytically explored. Specifically, we studied the behavior of the SD of MPH as a function of *H* and *T*. Figure 2 shows this relationship – the horizontal axis reports *H*, the vertical axis SD, and the differently shaped dots represent different *T*s.

MPH instability and inflation of false-positive rates

The simulation work fully confirmed the expected increase in instability of MPH with decreasing overall Hit rate – as visible in Figure 2, SD changes across two orders of magnitude. One example will explain why this happens, and contextually, will clarify why this leads to dramatic increases of false-positive rates (FPRs). Suppose that a neurologically intact subject who is severely long-sighted performs Weintraub and Mesulam's (1985) letter cancellation without wearing his glasses. On this task, one has to spot 60 small 'A' letters interspersed among 4–5 times more frequent distractor letters – a visually demanding task. His vision is so blurred that he detects only one target. Of course his MPH will be the position of that very target – the mean of a single Hit's position. The subject does not have neglect, so this target can occur with equal probability anywhere in the display – hence the very wide variability of MPH. Since the normal range covers 7.3% of the display width (data from Rorden & Karnath, 2010)[11], if we ideally repeated the experiment with the long-sighted subject many times, his only Hit (i.e. his MPH) would fall with 7.3% probability in the normal range, and with 92.7% probability outside it, leading to a false (left or right) neglect diagnosis – a disastrous FPR. The problem does not just affect the liminal case of 1 Hit, but any performance with suboptimal Hit rate – FPR would be 92.7% for 1 Hit, then it would decrease with increasing number of Hits, till it would reach about 2% (the nominal FPR) when Hit number is as high as in control subjects (wearing their glasses).[12] One can have a grasp of the size of the problem by looking at the top-left corner of Figure 2: those values mean that MPH has a SD of about .163 (out of an overall display width standardized at 1) when a patient without neglect detects 3/10 targets (white diamond). By applying a classical cut-off obtained from a control sample (we can use the .073-wide normal range obtained from Rorden & Karnath's, 2010 standardization) the .163 SD corresponds to an 82.2% FPR. Even taking less extreme examples, non-neglect patients canceling 25/50 targets (gray triangles) would have so variable MPHs as to be falsely classified as neglect with 39% instead of the nominal 2% probability. To use the exact target numerosity of the test from which the normal range was derived (Weintraub & Mesulam, 1985) non-neglect patients canceling 30/60 targets would have a FPR of 34.3%.

To summarize, MPH's inherent instability produces serious inflation of FPR. To our knowledge this factor, which equally affects any other neglect index, has never been taken into account. The problem is present whenever a patient misses targets for reasons that are different from neglect: any impairment, of any nature, decreasing detection probability will fatally increase the variability of MPH, thus producing false diagnoses in the absence of true neglect. Of course patients with brain damage are very likely to suffer from non-lateral deficits that reduce detection probability (we already listed some examples like amblyopia, agnosia, optic ataxia, dyslexia, low levels of concentration/motivation, disexecutive syndrome, etc.).

Ironically, we had chosen MPH as an optimal index because it is insensitive to the confounding effects of non-lateral deficits. Indeed, its long-run *average* value enjoys exactly this characteristic, however the *variance* of MPH strongly depends on non-lateral deficits and this is the cause of the problem.

Solutions to the problem of FPR inflation

Excluding all patients with non-lateral deficits from any procedure of neglect diagnosis is not a solution: this would leave no patient at all, especially if the neuropsychological assessment of functions other than neglect were exhaustive. Another idea is to recruit a group of brain-damaged patients without neglect, presumably suffering from non-lateral deficits, as a standardization sample. We think this cannot work either, for a number of reasons. First, non-neglect patients would indeed suffer from some non-lateral deficits, but these would be much less severe than in neglect patients, who typically have much larger lesions. Second, one would need to classify control patients as non-neglect in the first place: this would cause either a regression at infinity or a circularity.[13] Third, there would be inflation of FPR (for low Hit rates) or false-negative rate, FNR (for high Hit rates) because the assessed patient would be compared not with a sample of patients with exactly his/her level of non-lateral deficits, but with an unstructured sample of patients carrying a sizeable variability in this respect. Fourth, some subclinical cases of neglect would almost certainly sneak in the 'non-neglect' control sample,[14] bringing some extra variability, with further increase in FNR.

We think there is no *empirical/experimental* way out of this problem. Quite simply, the perfect control subject who suffers from exactly the same deficits as the patient to be assessed, but who does not have neglect, does not exist. Yet the problem must be solved otherwise all neglect tests would carry on suffering from inflated FPR. Remember that the same problem affects all neglect indices we are aware of.

We believe a straightforward *theoretical* solution exists. In short, the assessed patient is not to be compared to empirical control samples of any sort: his/her performance is to be contrasted against the null hypothesis that the Hits s/he produced – exactly that number of Hits – were generated with identical detection probability in all positions of the display – in other words, by non-lateral deficits. Thus, each patient is compared with a theoretical distribution and not against an empirical control sample. This is what we did in the next section.

Taking MPH instability into account: a correct diagnostic procedure

The MPH of the assessed patient is to be compared to the distribution of MPHs that s/he would have obtained, if s/he had had no neglect, but some non-lateral deficit that yielded his/her observed Hit rate. This is an easy task: that distribution has a zero mean and a SD that can be estimated from the data of our Monte Carlo simulations. For example, if the patient detected 20 out of 100 targets, the reference distribution would have a SD of about .06 (Figure 2, gray circles). By dividing the real, observed MPH by that SD one obtains an ordinary *z*-score to be used for diagnosing neglect. Such *z*-scores take into account the intrinsic instability of MPH and lead to correct *p*-values (i.e. to conventional 5% or 2% or any chosen α level) FPR. In other words, by following this procedure, the inflation of FPR disappears.

Of course we had to fit the clouds of points obtained from the simulations (Figure 2) by means of a general mathematical model, which gives a SD estimate as a function of *T* and *H*. In a complex mathematical work (detailed in the Website Material) we could find an Equation (see Appendix) that approximates SD values very satisfactorily (see solid curves in Figure 2), across the full ranges of studied *T* (from 10 to 150)[15] and *H* (from 1 to T–1). The Equation also adjusts the SD for the effect of target grouping (e.g. 50 targets in 10 instead of 50 positions). Therefore, by plugging *T* and *H* in the Equation one obtains a very precise estimate of SD and hence, a very precise estimate of a *z*-score that is free from FPR inflation. All mathematical complications can be overcome using an electronic Worksheet we developed (see later).

Normal subjects only vary for Hit rate and have no lateral biases (Diller task)

Our diagnostic procedure gave up neurologically intact subjects entirely. This was done under the assumption that all normal subjects have zero lateral bias, without inter-individual variation. In graphical terms, all normal subjects would have perfectly flat functions – isoprobabilities across space (like patient F, Figure 1). Inter-individual differences would only be allowed in the height of the isoprobability. This seemingly strong assumption does not have a serious impact on diagnosis error rates – especially if compared with the huge impact by non-lateral deficits in brain-damaged patients that was discussed above. However, we wish to consider such an assumption in some more detail, and to test it empirically. The assumption does *not* state that MPHs obtained from normal subjects should all be zero, indeed the observed MPHs depart from zero even though there is no lateral bias at all, just because subjects whose Hit rate is imperfect (<1) produce some sparse Omissions which 'move' MPH some tiny distance either way. The assumption states that *all* of the variation in the observed MPHs is due to those imperfect Hit rates, without any contribution by true lateral biases (tilts in the curves in Figure 1). We obtained an empirical proof that such an assumption is correct. The hypothesis that normal subjects do not have lateral biases predicts that the *z*-scores computed for them by means of our Equation should distribute with mean = 0 and standard deviation = 1 (taking the *z*-score from each subject is a way of partialling out inter-individual differences in Hit rate). We collected a sample of 199 controls (Female 57%, age 60.9 ± 12.2, education 11 ± 4.6) performing variants of Diller and Weinberg's (1977) letter cancellation task (targets ranged 104–108, were either 'H's or 'V's, and were administered on A4 or A3 sheets).[16] $N = 134$ subjects were excluded because of perfect performance (this leads to an unknown *z*-score, 0/0). The distribution of the remaining 65 subjects' *z*-scores almost perfectly matched the standard Gaussian: mean = –.024 (not significantly different from zero: one-sample *t*-test, $t(64) = .181$, $p = .857$, Bayes Factor = 1083 against the hypothesis that normal subjects lie .5 SD of 'pure noise' away from $z = 0$) and standard deviation = 1.077 (not significantly different from 1: $\chi^2(64) = 74.24$, one-tailed $p = .179$; Bayes Factor against hypothesis H_1 that variance = 2: BF = 18.675).[17] Therefore, the hypothesis that no normal subject has any true lateral bias was confirmed. The implication is that all of the variation in normal subjects' *observed MPH*s is the effect of the intra-individual noise caused by imperfect (<1) Hit rates. This is a direct confirmation of the validity of our statistical model, and of the neglect diagnoses yielded by it. Clearly such conclusions only hold for the Diller cancellation task, and we plan anyway to include in our Equation/Worksheet a term that accounts for possible inter-individual differences in lateral bias. However, the contribution of the (likely small) lateral biases of normal subjects to the variance of MPH is certainly minuscule – orders of magnitude smaller than the giant contribution of the intra-individual effects we found in our Monte Carlo study (Figure 2) and modeled with our Equation. According to it, a patient without neglect who detects 25/50 targets has expected MPH variance = .0018 which is 254 times larger than a plausible inter-individual variance in lateral bias of normal subjects (we considered a subject with 100% detection probability at one display end and 97% at the opposite end as having $|z| = 1$, which corresponds to inter-individual variation amounting to SD = .264% of the display width). So the current method is likely to capture virtually all of the variance relevant to diagnosis.

Software for automatic computation

We developed a Worksheet called 'MPH neglect diagnosis'. We are constantly updating the program, adding new functions, parameters, and making it more user-friendly. We also plan (May 2017) to link the worksheet to a graphic application allowing more immediate data insertion. The latest version of the program is available at psicologia.unipv.it/toraldo/mean-position-of-hits.htm. Alternatively, readers can email the authors (alessio.toraldo@unipv.it, alessio.toraldo@gmail.com, cristian.romaniello88@gmail.com, paolosom@gmail.com) or run a web search (use 'Measuring neglect' or 'unilateral neglect' and 'Mean Position of Hits' or 'MPH' as keywords). The current version of the Worksheet (May 2017) requires the user to insert the horizontal (and/or vertical) coordinate of each target and a '1' (Hit) or a '0' (Omission). The task that generated the data (visual, auditory, with hand movements, eye movements, or verbal responses, with simultaneous or successive presentation of each stimulus, with or without distractors, etc.) is entirely irrelevant, results will be valid anyway. Target number can range from 10 to 256. The output of the software is the MPH, for either dimension (horizontal/vertical), and all parameters relevant to diagnosis: standard deviation (SD), *z*-score, and *p*-value. Several other technical parameters are given, whose meaning is exposed in the Website Material.

MPH is given in three versions. The first is the MPH in the original scale (the one of the inserted coordinates). The second is the C-adjusted MPH ('C' for 'Centre') – in this standardized scale 0 corresponds to the mean position of all targets (MPT) and the display size (i.e. the distance between the left-most and right-most targets) is 1. C-MPH corresponds to what Rorden and Karnath (2010) proposed for cancellation (however they scaled the display size to 2). The third is the LCR-adjusted MPH ('LCR' means 'Left-Centre-Right') – in this continuous scale 0 corresponds to MPT, −.5 to the left-most target, +.5 to the right-most target. Each index is to be used for specific purposes. The C-version is slightly better for *diagnosing* neglect – it shows marginally better statistical behavior in liminal cases (the Worksheet automatically selects the C-version for computing *z*-scores and *p*-values). By contrast the LCR-version is better for *quantifying* neglect, because its −.5 and + .5 limits are stable anchors (in the C-version they can move a bit). In any case, if target distribution is reasonably homogeneous the two versions are virtually identical.

Note that the statistical model implemented in the Worksheet which outputs SDs and *z*-scores *assumes* that target distribution is perfectly homogeneous (i.e. with equispaced positions and equal number of targets per position). However, small departures from perfect homogeneity produce minuscule effects, which become completely negligible if one uses LCR-adjusted MPH. All neglect tasks we know use very homogeneous fields of targets (deliberately so, in order to avoid polluting the effects of a genuine bias in the patient's processing system with a bias in distribution of stimuli), so our statistical model works very well for them. Nonetheless, the Worksheet includes various diagnostics for severe departures from perfect homogeneity: a warning message is given, for example, when targets are so left–right imbalanced that their mean position, MPT, is more than 5% of the overall display size away from the display geometrical center. In these cases MPH is not given, and, if appropriate (see Website Material for details) the user is advised to look at a 'non-parametric' version of the MPH, the MOH (Mean Ordinal position of Hits). While MPH is the mean position of Hits in the physical space of the display, the MOH is the mean position of Hits in the abstract space of ordinal positions of the targets across the display. Clearly, any statistical inference is only valid in this abstract ordinal space if MOH is used.

General discussion

We wished to address a widespread problem that both clinical and experimental neuropsychologists face when diagnosing neglect, that is, the large variability of tasks used and the absence of a standard statistical procedure. There is room for an attempt towards uniformity, since most neglect tasks share the same data form: they require processing of stimuli – targets – that are administered in different positions, and with each target receiving a Hit/Miss dichotomous score. So our question was whether an *a priori* justified index and a valid statistical model exist for quantifying and diagnosing neglect on such tasks. Clinicians have almost always used simple measures, like the difference in Hit counts or rates between the two halves of the tested space, or the overall Hit Rate, because of their immediate comprehensibility. However, we showed how such measures fail to discriminate between cases with obviously different neglect severities (Table 1).

In search for a new index, we preferred to follow pure theoretical reasoning rather than an in-depth analysis of the literature, in the hope of formulating a score with more general validity. Thus, we analyzed a systematic series of 'thought experiments', groups of virtual patients whose performance curves differ for a single mathematical parameter (Figure 1). A thorough theoretical discussion of these cases led to the Mean Position of Hits (detected targets), MPH, as a good candidate measure since it correctly orders patients for neglect severity and also has nice close-to-Gaussian distributions. Most importantly MPH partials out the confounding effects of concomitant *non-lateral* deficits on performance; these deficits, which are different from neglect, decrease the Hit rate by a constant factor through space, and MPH is – on the average – invariant to this interference. Therefore, we would encourage the use of MPH as a theoretically meaningful neglect index, valid for a vast family of screening and research tasks.

In spite of these desirable characteristics, MPH suffers from a problem that actually affects all neglect indices (including the simple ones, left–right difference in Hit rate, overall *Hit Rate*). While the *average MPH* is immune to the effects of concomitant non-lateral deficits, the *variance* of MPH is not: if a patient has non-lateral deficits, these increase the SD of MPH up to two orders of magnitude (Figure 2), leading to serious inflation of the rate of false neglect diagnoses. Because brain-damaged patients can be, and almost always are, affected by non-lateral deficits this problem is likely to have sizeable practical consequences: many patients who do not have neglect but suffer from severe, different cognitive deficits are classified and rehabilitated as if they had neglect. It is rather surprising that this problem has been (to our knowledge) entirely overlooked in the literature. For example, we estimated the FPR to be 39% for non-neglect patients detecting 25/50 targets.

The solution we proposed is to take into account the instability of MPH in the diagnostic procedure (this instability could be estimated, and precisely modeled, in a thorough statistical-mathematical study with Monte Carlo simulations, see Figure 2 and Website Material) by computing a *z*-score for each patient exactly on grounds of such instability. This technique automatically eliminates all inflations of FPR, providing a statistically correct diagnosis of neglect.

The estimation of degree of instability (SD) of MPH is painstakingly complex (see Appendix), so we also provided downloadable software that allows one to obtain correct neglect diagnoses (MPH, expected SD, *z*-score, and *p*-value) directly from raw data. The interested reader can find relevant information in the section 'Software for automatic computation' or contact the authors directly.

Although still not widespread, MPH has already been used in some neglect research, especially in cancellation (Binder et al., 1992; Dalmaijer, Van der Stigchel, Nijboer, Cornelissen, & Husain, 2015; Gandola et al., 2007; Rorden & Karnath, 2010; Gandola et al., 2013). Rorden and Karnath (2010) highlighted the advantages of CoC ('Centre of Cancellation', the C-MPH in a cancellation task) on grounds of a review of the indices proposed in the previous literature. What we add to Rorden and Karnath's analysis is (i) the theoretical analysis yielding *a priori* justification for the use of MPH, (ii) the statistical model protecting the clinician/researcher from the risk of inflated FPR, and (iii) the empirical demonstration that normal controls are free from lateral biases (at least in one wide-spread cancellation task) which allows one to assume that all of normal MPH variability is due to the effects of sub-optimal, but spatially invariant, Hit probability.

By 'a-priori justification' we refer to the fact that we derived the MPH from a theoretically-driven set of thought experiments, rather than from an in-depth analysis of the literature. This difference is relevant because our theoretical architecture clarifies *why* the MPH is a good measure, and in *what exact conditions* it is so. Thus researchers who support a specific theory of neglect – say, the attentional theory (Posner et al., 1984) can directly verify whether or not that theory agrees on the classification of neglect severities in our virtual patients (Figure 1). If it agrees then MPH will be a justified neglect index, if it does not agree MPH will be an unreliable measure. Therefore, we offered a way of bridging the gap between the debate on the validity of indices and the theoretical debate on the nature of neglect.

Limits

The limits of this study are in the assumptions we made for performing the simulations (the Website Material reports a full list). Clearly we could not simulate all of the possible target distributions – virtually infinite in number. On the other hand, we did not wish to limit our analysis to just a few exact target distributions (i.e. to some real tasks) because we aimed at providing a tool that can be used with *every* task characterized by the 'Hit/Miss-by-position' data form. Thus, the deal was to simulate data from ideal, perfectly homogeneous distributions, considering that distributions in real tasks are very close to this situation, and including non-homogeneity diagnostics in our Worksheet.

Another limit is in the assumption that *all normal subjects have zero lateral bias*. We mean that, while detection probability can differ in different subjects, in each and every of them it is assumed to be constant across space. We confirmed this assumption to be true in a sample of 65 subjects performing the Diller and Weinberg (1977) cancellation test. However, even if the assumption did not hold for other tasks, normal subjects are likely to show minuscule biases, literally orders of magnitude smaller than the effects captured by our model (i.e. those caused by sub-optimal Hit rates, which especially affect the performance of brain-damaged patients). Hence, our Equation is certainly robust to the presence of tiny differences in lateral bias within non-neurological samples. In any case, we plan an extension of our model towards inclusion of such (however small) effects. The interested reader can follow the development of the computerized program on the Website.

Another assumption reduces the applicability of the present technique in some specific tasks. Using the terminology of the attentional theory of neglect (Posner et al., 1984), this assumption states that *if a target is not reached by the attentional focus, a Hit is impossible*.[18] Most tasks obey this assumption. For instance, if a patient does not shift his/her attentional focus onto a target a reaching (or eye) movement towards it will never be made, so all tasks requiring those kinds of response are safe. However, there are a few types of task in which a Hit is indeed possible even without attention. For example, in classical psychophysical paradigms visual stimuli are delivered one at a time together with a 'prompt' signal (e.g. a 'beep' sound) and subjects have to say 'yes, I saw it' or 'no, I did not see it'. It is well possible that subjects guess a 'yes' response after the beep even though they failed to allocate attention over a stimulus (i.e. without 'seeing' it). Such tasks typically have catch trials in which no stimulus is delivered, so one can directly test whether the assumption was violated (in this case 'yes' responses, i.e. False Alarms, would be observed in catch trials), or not (in this case FA would not be observed). Violations of the assumption cause underestimation of the absolute MPH (the more severe, the higher the FA rate: e.g. if FA rate = .2, MPH is underestimated by 33.5%; if FA rate = .4, by 57.3%) and consequent reduction in diagnostic sensitivity. A general classification of tasks with respect to this problem is available in the Website Material; in short, the critical (rare) tasks are those in which a single target is delivered on each trial, response is symbolic (e.g. vocal response or button press), it is over very few categories (e.g. yes/no), and is prompted on each trial. Future work will address this issue.

Generalizability to other syndromes (or to other research fields)

We wondered whether the present analysis might be applied in the evaluation of other syndromes. Obvious candidates are hemianopia and hemianesthesia: in both cases stimuli are presented along a continuum of physical positions (e.g. static perimetry for the visual case) and each response can be classified as a Hit/Miss. However consider that, while our model assumes that normal subjects have isoprobability of detection across the studied perceptual space, this might not hold for brief stimuli that are administered in different retinal or somatic positions. Clearly visual detection probability decreases with retinal eccentricity and tactile sensitivity varies according to the stimulated somatic region. The optimal strategy would be that of adjusting stimulus intensity in order to equalize sensitivity all across the tested sensory space (in normal subjects). If this were not possible, users of our MPH analysis should consider that a physiological decrease (increase) of Hit rate with eccentricity in the sensory space causes inflation of the false-negative (-positive) rate in the diagnosis of a lateral deficit.[19]

One last remark is that the mathematical/statistical machinery proposed here can be applied not only to distributions of Hits/Misses across spatial positions, but to distributions of any type of dichotomous scores across any type of continuum (e.g. word frequency, sound pitch, earthquake intensity, etc.), provided that all the assumptions of the model hold.

Conclusions

We propose a statistical model which helps standardizing the diagnostic procedure of a very large number of different neglect tests by exploiting their basic mathematical similarity. We believe the most valuable features of the present MPH analysis/software are: (i) MPH 'extracts' the neglect component from the background of a multitude of concomitant, non-lateral deficits and (ii) the MPH statistical model avoids the inflation of false-positive rates that any classical standardization based on a sample of control subjects would unavoidably yield.

Notes

1. We often use 'detection' instead of 'successful processing' in the paper, albeit 'detection' would be inappropriate in some instances where the 1/0 score dichotomy applies (e.g. when the task is recall from memory).
2. R-L compares the two halves of the display. One can divide space in more than two sectors and compute a regression of % Hits against the sectors' spatial positions; in the limit one may take the Hit/Miss score on each single target, plot it against position, and fit the cloud of points with something like a logistic curve. Butler, Eskes, and Vandorpe (2004) did so and quantified neglect in terms of the *slope* of the curve. However exactly the same logical limits affecting the simple *R–L* difference (Table 1) also affect regression *slope*. See later for further discussion of this parameter.

3. By this expression we mean that either sheer intuition or more formalized models of neglect indicate a specific order of neglect severity. We wished not to commit to any general or specific theory of neglect in the present paper in order to keep the validity of our analysis as general as possible.
4. *Location* varies in the domain ±∞, not in the (−.5, .5) space of the display; *ceiling* is the top hit rate in the ±∞ domain, or the upper asymptote of the curve; *slope* is relative to *ceiling*: it means how 'fast' the Hit rate drops towards zero starting from *ceiling* (not from 1).
5. Also *ceiling* is constant. Recall that *ceiling* is the upper asymptote: at +∞, it reaches 1 for all three patients.
6. This is a reasonable assumption, since double dissociations between neglect and many other deficits potentially affecting target detection have been shown. About terminology, if one is studying the vertical dimension, the term 'non-altitudinal' would be more correct than 'non-lateral'. However we use the term 'non-lateral' all across the paper for clarity's sake.
7. Suppose that on a cancellation task with distractors, in position *x* the level of neglect (the 'lateral' deficit) is such that 50% of stimuli are missed. Suppose also that because of mild visual agnosia (the 'non-lateral' deficit) the patient correctly processes the shape of a target with 70% probability. The final Hit rate in position *x* is .5 × .7 = .35 – one has *both* to spatially select ($p = .5$) *and* to correctly process the shape ($p = .7$) of a target in order to cancel it. By changing position, the probability of successful spatial parsing changes (it increases rightwards and decreases leftwards) while the probability of successful shape processing is constant across space, $p = .7$.
8. If one accepted the idea that G-I profiles can reflect different degrees of neglect, the differences been those profiles would become entirely ambiguous: they might reflect differences in neglect severity, differences in non-lateral deficits, or differences in both. Interpretation of performances D-F would be affected by the same ambiguity.
9. The comparisons we carried out were between patients differing for *only one* parameter of the curve at a time. We did not discuss the (virtually infinite) cases where differences in more than one parameter combine: their complexity would have made any decision as to the correct order of neglect severity impossible without an explicit neglect theory, and we wished to avoid committing to any such a theory, in order to keep our statistical model as general as possible. Hence we were satisfied that the chosen index, MPH, could correctly differentiate the equivalence classes obtained from variation of single parameters of the curve.
10. The instability of *R–L difference* and *Accuracy* is maximal when (average) Hit rate = .5 and minimal when it approaches 0 or 1. L/TOT's instability, like MPH's, is minimal for (average) Hit rate close to 1 and maximal when it approaches 0.
11. Rorden and Karnath (2010) obtained cut-offs for CoC from 53 control patients without neglect. The sample's standard deviation was .0313, hence a normal range with bidirectional FPR = 2% is .146 wide. However the CoC scale is (−1, 1) so in terms of the MPH scale (−.5, .5) the normal range covers 7.3% of the display width.
12. We have a dramatic, empirical example of a false positive. One of the 199 *neurologically intact* subjects who performed Diller and Weinberg's (1977) cancellation task (see the section 'Normal subjects only vary for Hit rate …') had an exceptionally low Hit rate, .722. Taking as a 'standardization sample' the 197 subjects whose detection rate was at least .9, his *z*-score for the MPH was +5.389 which represents strong evidence of left neglect. Our method, which takes into account the large increase in MPH variance due to decrease in overall detection rate, led to a *z*-score of .703 – perfectly within normal variability.
13. Call the test(s) used for filtering the patients of the control group 'F' (for 'Filtering'); call the test to be standardized 'ST'. If *F* and ST are the same test, the argument is circular: one would be using ST to select control subjects that are needed to standardize ST. Also if *F* is just *similar* (correlated) to ST the argument is circular: one would be using variation in *F*, that is shared by ST, to select subjects needed to standardize ST. Else, if *F* is completely different from (uncorrelated with) ST, the argument is a regression at infinity: *F* was used in the process of standardization of ST, but in order for *F* to be valid, also *F* should be standardized, so another, different test F_1 should be used to standardize *F*, and so on.

14. The only truly safe way of excluding subclinical neglect is to exclude all patients with brain damage, that is, to use only normal subjects.
15. The Equation works perfectly well also for $T = 256$ – the upper limit of targets in the Worksheet for automatic computation.
16. Data were collected from the electronic archives of many different experimental and clinical studies carried out by one of us (AT) across several years (1994–2013). Demographics could be traced back for 76% of the subjects.
17. There was a clear outlier, with $z = 3.318$. This subject missed 11/108 targets, 10 on the left and 1 on the right display half. The absolute deviation of his MPH was minor (+.03 or 3% of the display width), however even excluding him on suspicion of some undetected minor brain damage (a legitimate move: H_1 specifies $\sigma > 1$ with Gaussian shape, and not that there are outliers) group mean = −.076 ($t(63) = .611$, $p = .543$, BF = 9210) and standard deviation = .999 ($\chi^2(63) = 62.9$, one-tailed $p = .48$, BF = 225.12).
18. In more general terms: *if the processes that induce neglect when lesioned, fail to parse a given target, a Hit is impossible.*
19. If eccentricity effects are markedly left-right asymmetrical, this would cause inflation of both FPR and FNR.

Acknowledgments

We thank Robin Hilsabeck and the two anonymous referees for their thoughtful comments, and Justin Harris for his kind help in revising the English. This paper is dedicated to the memory of Emanuela Radice and Alessandro Latocca.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Albert, M. (1973). A simple test of visual neglect. *Neurology, 23*, 658–658. doi:10.1212/WNL.23.6.658

Binder, J. R., Marshall, R., Lazar, R. M., Benjamin, J., & Mohr, J. P. (1992). Distinct syndromes of hemineglect. *Archives of Neurology, 49*, 1187–1194. doi:10.1001/archneur.1992.00530350109026

Bisiach, E., Capitani, E., & Porta, E. (1985). Two basic properties of space representation in the brain: evidence from unilateral neglect. *Journal of Neurology, Neurosurgery, and Psychiatry, 48*, 141–144. doi:10.1136/jnnp.48.2.141

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex, 14*, 129–133. doi:10.1016/S0010-9452(78)80016-1

Bottini, G., Gandola, M., Scarpa, P., Toraldo, A., & Zanardi, G. (2005, June). *Productive symptoms in the personal and peripersonal space associated with neglect*. Presented at the International Neuropsychological Symposium, Alghero, Italy.

Butler, B. C., Eskes, G. A., & Vandorpe, R. A. (2004). Gradients of detection in neglect: Comparison of peripersonal and extrapersonal space. *Neuropsychologia, 42*, 346–358. doi:10.1016/j.neuropsychologia.2003.08.008

Dalmaijer, E. S., Van der Stigchel, S., Nijboer, T. C., Cornelissen, T. H., & Husain, M. (2015). CancellationTools: All-in-one software for administration and analysis of cancellation tasks. *Behavior Research Methods, 47*, 1065–1075. doi:10.3758/s13428-014-0522-7

De Renzi, E., Gentilini, M., & Barbieri, C. (1989). Auditory neglect. *Journal of Neurology, Neurosurgery, and Psychiatry, 52*, 613–617. doi:10.1136/jnnp.52.5.613

Diller, L., & Weinberg, J. (1977). Hemi-inattention in rehabilitation: The evolution of a rational remediation program. In E. A. Weinstein & R. P. Friedland (Eds.), *Hemi-inattention and hemisphere specialization* (pp. 63–82). New York, NY: Raven Press.

Edgworth, J., Robertson, I. H., & MacMillan, T. (1998). *The Balloons test: A screening test for visual inattention*. Bury St Edmunds: Thames Valley Test Company.
Friedman, P. J. (1992). The star cancellation test in acute stroke. *Clinical Rehabilitation, 6*, 23–30.
Fruhmann-Berger, M., & Karnath, H. O. (2005). Spontaneous eye and head position in patients with spatial neglect. *Journal of Neurology, 252*, 1194–1200. doi:10.1007/s00415-005-0831-y
Gandola, M., Toraldo, A., Invernizzi, P., Corrado, L., Sberna, M., Santilli, I., … Paulesu, E. (2013). How many forms of perseveration? Evidence from cancellation tasks in right hemisphere patients. *Neuropsychologia, 51*, 2960–2975. doi:10.1016/j.neuropsychologia.2013.10.023
Gandola, M., Toraldo, A., Paulesu, E., Corrado, L., Invernizzi, P., & Bottini, G. (2007, October). *Anatomical correlates of different types of perseveration in patients with neglect*. Poster presented at the Body Representation Workshop, Rovereto, Italy.
Gauthier, L., Dehaut, F., & Joanette, Y. (1989). The Bells test: A quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology, 11*, 49–53. Retrieved from http://strokengine.ca/assess/pdf/bellstest.pdf
Halligan, P., Robertson, I. H., Pizzamiglio, L., Hömberg, V., Weber, E., & Bergego, C. (1991). The laterality of visual neglect after right hemisphere damage. *Neuropsychological Rehabilitation, 1*, 281–301. doi:10.1080/09602019108402259
Karnath, H. O. (1997). Spatial orientation and the representation of space with parietal lobe lesions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 352*, 1411–1419. doi:10.1098/rstb.1997.0127
Moreh, E., Malkinson, T. S., Zohary, E., & Soroker, N. (2014). Visual memory in unilateral spatial neglect: Immediate recall versus delayed recognition. *Journal of Cognitive Neuroscience, 26*, 2155–2170. doi:10.1162/jocn_a_00603
Pitteri, M., Venneri, A., Meneghello, F., & Priftis, K. (2013). How to differentiate hemianesthesia from left tactile neglect: A preliminary case report. *Behavioural Neurology, 26*, 151–155. doi:10.3233/BEN-2012-110225
Posner, M., Walker, J., Friedrich, F., & Rafal, R. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience, 4*, 1863–1874. Retrieved from http://www.jneurosci.org/content/4/7/1863.long
Rorden, C., & Karnath, H. O. (2010). A simple measure of neglect severity. *Neuropsychologia, 48*, 2758–2763. doi:10.1016/j.neuropsychologia.2010.04.018
Toraldo, A., Laiacona, M., Pagani, R., Mandelli, A., & Capitani, E. (2014). Perceptual and response-related components of unilateral neglect may evolve independently of one another: Evidence from five single-case studies. *Neurocase, 20*, 241–259. doi:10.1080/13554794.2012.741257
Toraldo, A., McIntosh, R. D., Dijkerman, H. C., & Milner, A. D. (2004). A revised method for analysing neglect using the landmark task. *Cortex, 40*, 415–431. doi:10.1016/S0010-9452(08)70136-9
Weintraub, S., & Mesulam, M. (1985). Mental state assessment of young and elderly adults in behavioral neurology. In M.-M. Mesulam (Ed.), *Principles of behavioral neurology* (pp. 71–123). Philadelphia, PA: F.A. Davis Company.

Appendix

The final model's equation for the SD of MPH follows (see Website Material for explanations).

$$\sigma_{\mathrm{MPH}} = (\mathrm{CF})[(150^2 - 1)/(3H)]^{1/2} T(1 - W)/[300(T - 1)]$$

In it: T = overall number of targets; H = number of Hits by the patient; $W = j(H/T)^3 + k(H/T)^2 + m(H/T) - (j + 150\,k + 150^2\,m)/150^3$ where scalar values are j = 1.155, k = –.56, m = .56; CF = $(.5/D - q - 1)x^2 + qx + 1$ where $D = [(T^2 - 1)/12]^{1/2}/(T - 1)$, scalar q = .477, $x = 2(T - G)/[G(T - 2)]$, and G within x is the number of clusters in which targets can best be grouped. G is estimated on grounds of the empirical target distribution: one identifies G equispaced positions ('knots') starting from the leftmost and ending at the rightmost target position, and computes the square distances of each target from its closest knot; this operation is repeated for all Gs between 2 and T, and the smallest G among those that minimize the sum of the square distances is selected. If the distribution of targets is not homogeneous across the tested space, G estimation is unreliable. The

Equation provides precise SD estimates provided that $T > 9$. p-values can be obtained by Gaussian approximation [$z = $ (C-adjusted MPH)/SD] when there are at least 3 Hits and 3 Omissions [$2 < H < (T-2)$, finer criteria are reported in the Website Material], otherwise our Worksheet provides non-parametric solutions.