BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Additive scales in degenerative disease - calculation of effect sizes and clinical judgment

Matthias W Riepe[1*], David Wilkinson[2], Hans Förstl[3] and Andreas Brieden[4]

Abstract

Background: The therapeutic efficacy of an intervention is often assessed in clinical trials by scales measuring multiple diverse activities that are added to produce a cumulative global score. Medical communities and health care systems subsequently use these data to calculate pooled effect sizes to compare treatments. This is done because major doubt has been cast over the clinical relevance of statistically significant findings relying on *p* values with the potential to report chance findings. Hence in an aim to overcome this pooling the results of clinical studies into a meta-analyses with a statistical calculus has been assumed to be a more definitive way of deciding of efficacy.

Methods: We simulate the therapeutic effects as measured with additive scales in patient cohorts with different disease severity and assess the limitations of an effect size calculation of additive scales which are proven mathematically.

Results: We demonstrate that the major problem, which cannot be overcome by current numerical methods, is the complex nature and neurobiological foundation of clinical psychiatric endpoints in particular and additive scales in general. This is particularly relevant for endpoints used in dementia research. 'Cognition' is composed of functions such as memory, attention, orientation and many more. These individual functions decline in varied and non-linear ways. Here we demonstrate that with progressive diseases cumulative values from multidimensional scales are subject to distortion by the limitations of the additive scale. The non-linearity of the decline of function impedes the calculation of effect sizes based on cumulative values from these multidimensional scales.

Conclusions: Statistical analysis needs to be guided by boundaries of the biological condition. Alternatively, we suggest a different approach avoiding the error imposed by over-analysis of cumulative global scores from additive scales.

Keywords: dementia, neurodegeneration, clinical studies, meta-analysis, effect sizes, Cohen's d

Background

Analysis of treatment efficacy is warranted to guarantee the quality of medical treatment and effective spending of resources. Across diseases, meta-analyses are assumed to be one the major tools to achieve this [1-4]. Meta-analyses are performed to come to an overall conclusion on clinical studies with different numerical results or using different assessment methods. One critical step in performing meta-analyses is to calculate the effect sizes for the studies to be included in the meta-analysis [5].

Degenerative diseases are of long duration and the diversity of their symptoms pose methodological difficulties not known in other fields of medicine: symptoms vary over time, fluctuate for random reasons, and may be replaced by new and different ones. To illustrate the reasoning on whether effect sizes and meta-analyses are suited to resolve the ambiguity of clinical study results in degenerative disease one of the most prevalent degenerative diseases, Alzheimer's disease (AD), will be used.

AD is the most frequent cause of dementia in old age and typifies the variability in clinical presentation and symptom changes over time that occurs in a degenerative disease. At onset of AD the medial temporal lobe is affected [6]. This results in the episodic memory deficit

* Correspondence: matthias.riepe@uni-ulm.de
[1]Department of Psychiatry and Psychotherapy II, Mental Health & Old Age Psychiatry, Ulm University, Ulm, Germany
Full list of author information is available at the end of the article

wihich is an early clinical hallmark of the disease [7]. As the disease spreads, other brain regions such as the frontal and parietal cortex are affected as well. The parietal cortex mediates activities such as spatial orientation and visuo-spatial functions [8,9]; the frontal cortex mediates executive functions, planning, attention, and working memory [10-12]. Spread of AD beyond the temporal lobe thus is characterized in functional terms by accruing deficits of spatial orientation, attention and executive functions as well as working memory and language [7]. This affliction of different brain regions and functions can be visualized using advanced imaging methods [13-15]. Despite an overall progress, symptoms may also fluctuate over the course of progressing dementia for random reasons. Apathy may turn to agitation which may disappear and followed by apathy, again. Regardless of this complexity, effect size calculation and meta-analyses of different studies use the addition of scores from many disparate functions to provide a global score for problems like 'cognition', 'behavior', or 'activities of daily living'. 'Cognition' comprises a multitude of activities such as episodic or working memory, attention, calculation, cognitive flexibility, praxis; 'behavior' comprises affect and emotion, delusion, agitation, irritability, and 'activities of daily living' comprise a wide variety of tasks for which the performance not only depends on the actual capabilities of the patient but also on her or his prior habits. Over the whole course of the disease, 'cognition' or 'behaviour' may be appropriate to assess overall dementia but over the time frame of clinical studies, usually one to two years, individual cognitive functions need to be focused on as the disease process over such short time spans is confined to specific functions and specific regions rather than the whole brain. At present, however, and for the last 30 years, clinical studies in AD have used global scales, i.e. multidimensional scales, to appraise the efficacy of interventions using instruments such as the Alzheimer's Disease Assessment Scale (ADAS) [16], the Mini-Mental-Status Examination (MMSE) [17], the Severe Impairment Battery (SIB) [18], the Neuropsychiatric Inventory (NPI) [19], the Katz activities of daily living scale (Katz-ADL) [20] amongst others.

Physicians and statisticians not well acquainted with the administration of neuropsychological tests neglect the impact of test difficulty on neurobiological associations. Task difficulty has a profound impact on the neural substrates engaged to solve the task. It was shown recently, that task difficulty is associated with recruitment of different neural patterns even in healthy subjects [21]. Thus, despite being similar activities, two tasks may rely on the integrity of different brain areas if the tasks vary in difficulty. Clearly then, the likelihood of maintaining performance on a specific task being measured with a particular instrument is dependent on disease severity and on time since diagnosis. The task may rely on different areas of the brain being recruited as degeneration reduces the relative amount of input from areas normally engaged in that function and showing a non-linear decline in dementia patients [22,23].

Multidemensional clinical scales combine different tasks, i.e. different activities, to assess overall severity of brain dysfunction. The cumulative score for these multidimensional scales results from summation of subscores representing specific activities. The relative contribution of the sub-scores to the total score, however, is variable, as is the task difficulty to assess specific activities in the different scales (e.g. the MMSE has a total score of 30 and scores 3 points for the recall of three words on single presentation and that task which is preserved till very late in the disease carries the same weight as the three points that could be obtained from recalling those words 5 minutes later a task that is very often one of the earliest signs of impairment, the ADAS-cog asks for recall of ten words on threefold presentation of the test and together with other memory items the function memory is represented with 27 points out of 70).

It was our goal to address the impact of non-linearity of disease progression and construction of multidimensional scales on the analysis of these additive global scales.

Methods

Basic model for the representation of function

Modeling the decline of function needs to reflect that tasks that are easy show a ceiling effect in assessment in early disease (i.e. the task is so easy or the underlying brain circuits are so insensitive to the disease process that the score does not decline over the initial time of the degenerative process) and in the later stages a floor effect (i.e. the task is so difficult or the underlying brain circuits are so severely affected from the disease process that the score is not sensitive enough to pick up further decline). Such a pattern was demonstrated for the items of the Mini-Mental-Status Examination [23,24], repeating of words is task with an early ceiling effect and delayed recall of memorized words is a task with an early floor effect. Accordingly we used an inverse exponential rule for modeling the decline of function with progressing disease: $f_i(t; a_i, b_i, c_i) = \left(a_i + b_i t e^{c_i/t}\right)^{-1}$, where $i = 1, 2$, $t_{min} \le t \le t_{max}$, $c_i < 0$.

Different f_i represent different symptoms (e.g. memory, praxis, and so forth) declining over time according to parameters a_i, b_i, and c_i, accessible by empirical

studies, and t indicating time. Qualitatively, the arguments outlined below are also valid for various other functions than the inverse exponential function.

Results

Vulnerability and difficulty

Two examples for the decline of performance over time using the basic model are shown in Figure 1.

These curves can be interpreted in two different ways: I) function f_1 and f_2 represent different tasks, e.g. memory and praxis. In this interpretation, f_1 represents an activity that early and rapidly declines with progression of disease (e.g. episodic memory in patients with Alzheimer's disease). The function f_2 represents an activity that is upheld early during progression of disease with decline only occurring later (e.g. praxis in patients with Alzheimer's disease). Within this framework the neurobiological reason for the distinct time course of decline of function is selective vulnerability of brain regions. II) Alternatively, it may be assumed that the two curves represent the same task (e.g. spatial orientation). With this interpretation f_1 represents measurement of the task with an instrument without a ceiling effect but with an early floor effect (e.g. spatial orientation in an unknown environment in patients with Alzheimer's Disease). The function f_2 in this interpretation represents an instrument with an early ceiling effect and a late floor effect (e.g. spatial orientation in a known environment in patients with Alzheimer's Disease). In other words, f_1 has a high task difficulty (reflecting disease progression or design of instrument) and f_2 has a low task difficulty (reflecting disease progression or design of instrument).

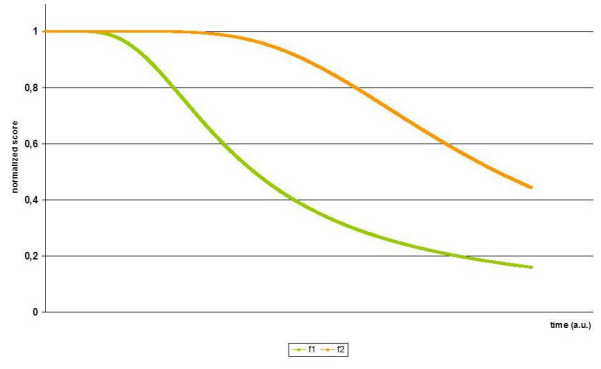

Figure 1 Selective vulnerability and task difficulty. $f_i(t; a_i, b_i, c_i) = \left(a_i + b_i t e^{c_i/t}\right)^{-1}$, i = 1, 2, $t_{min} \le t \le t_{max}$.For the orange curve the parameters of the formula in Figure 1 are: $a_1 = b_1 = 1$, $c_1 = -1/6$. For the green curve the parameters are: $a_2 = b_2 = 1$, $c_2 = -1/20$. The orange curve represents a symptom with a ceiling effect at the beginning of clinical disease (e.g. praxia in Alzheimer's disease), the green curve represents a symptom with a floor effect early during progression of disease (e.g. episodic memory in Alzheimer's disease).

Multidimensional additive scales

We now assume two scales (e.g. the MMSE and the ADAScog), one scale represented by F_A and another scale represented by F_B, both comprised of two tasks following functions f_1 (a task that declines early and rapidly over the course of disease) and f_2 (a task that declines later during the course of disease) but weighted differently in F_A and F_B:

F_j (t; a_i', b_i, c_i, λ_{ji}, i = 1, 2) = $\lambda_{j1} f_1$(t; a_1', b_1, c_1) + $\lambda_{j2} f_2$ (t; a_2', b_2, c_2) for $j \in$ {A, B} where λ_{j1} λ_{j2} with which the functions f_1 and f_2 are weighted in the scales F_A and F_B, respectively. Without loss of generality: $\lambda_{j1} + \lambda_{j2}$ = for $j \in$ {A, B}.

To illustrate it: the cognitive part of the Alzheimer's Disease Assessment Scale (ADAScog) weights 'memory' with 27 out of 70 points: (word recall (max. 10), word recognition (max. 12), remembering test instructions (max. 5)). The Severe Impairment Battery (SIB) weights 'memory' with a maximum of 14 out of 100 points. The Mini Mental State Examination weights 'memory' with 6 out of 30 points. In contrast, 'orientation' is reflected in these scales with a maximum of 8 out of 70, 6 out of 100, and 10 out of 30, respectively.

How combination of assessment of different tasks into one scale affects assessment of disease progression as measured with these scales is shown in Figure 2.

Treatment effects

We now assume treatment affects by scaling factors 1 + δ_i, i = 1, 2, such that reflecting a purely symptomatic treatment effect on the progression of the disease for the treated group is described as (1 +δ_i) f_i (t; a_i, b_i, c_i)

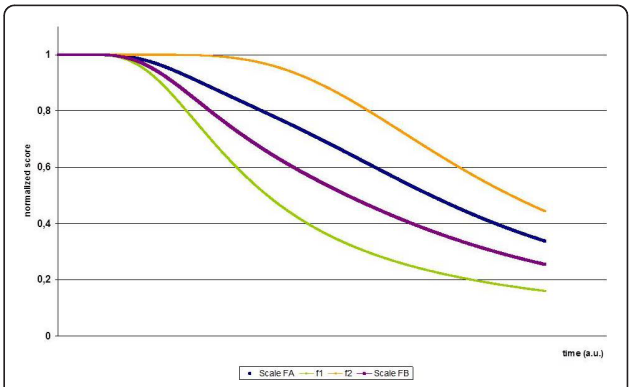

Figure 2 Composite Scales. Functions f_1 and f_2 as in Figure 1. Scale F_A: F_A(t; a_i, b_i, c_i, i = 1, 2) = 3/8 f_1(t; a_1, b_1, c_1) + 5/8 f_2 (t; a_2, b_2, c_2). Scale F_B: F_B (t; a_i, b_i, c_i, i = 1, 2) = 2/3 f_1(t; a_1, b_1, c_1) + 1/3 f_2 (t; a_2, b_2, c_2):. Hence, the scale F_A is dominated by function f2 and scale F_B is dominated by function f1. The graph shows normalized scores over time.

for $i = 1, 2$, $t_{min} \le t \le t_{max}$ Comparison of effect sizes or calculation of a common effect size in a meta-analysis naturally has to assume time-independence of the effect size - otherwise the result of bringing together results from multiple studies would strongly depend on how many studies with milder or more advanced severity of patients, respectively, are brought together in the analysis. The mathematical analysis below shows that a sufficient condition in the mathematical sense to achieve time independent effects is to assume that the standard deviation is proportional to the mean of the observed data. From a practical point of view this can be interpreted as a constant relative deviation. More precisely, Theorem 1 states that the effect size Cohen's d of both measurements is independent of the time of observation, i.e., $d_i(t) \equiv d_i$ Hence, the necessary condition for applying for applying meta-analysis is satisfied. However, in general meta-analyses can also be performed with cumulative values of multidimensional scales and the question of time-independent effects have to be answered again. For this consider the additive scales F_j $(t, a_i, b_i, c_i, \lambda_{ji}, i = 1, 2) = \lambda_{j1} f_1(t; a_1, b_1, c_1) + \lambda_{j2} f_2(t; a_2 b_2 c_2)$ for $j \in \{A, B\}$ introduced before. Time-independence would follow if the effect sizes needs to be calculated in the intuitive way as $d_j(t) = \lambda_{j1} d_1(t) + \lambda_{j2} d_2(t)$. "Unfortunately", mathematical analysis (see below for more details) yields in that the effects size is a function depending on the weights $\lambda_{j1}, \lambda_{j2}$ $j \in \{A, B\}$ of the functions f_1 and f_2 in the composite scales F_A and F_B, the treatment effects δ_1, δ_2, and in contrast to the intuition in general on the functions f_i, $i = 1, 2$, and - most important - the time t (Figure 3).

It is natural to ask, under which assumptions we can get rid of the general statement on time-dependence and still can guarantee time-independence for additive scales. The mathematical analyses shows that this is the case if we assume that over time the observed data are perfectly correlated with respect to the different scales and in addition if $\delta_1 = \delta_2$ (this means that the treatment effect is identical for both functions f_i, $i = 1, 2$, representing different cognitive functions) or $\lambda_i = 0$, $i \in \{1, 2\}$ The latter assumption means that function of interest is no longer multidimensional. Whether these assumption are either realistic or of relevant interest has to be decided in a preprocessing step.

However, in order to be able to calculate the time-dependent scaling factor in the general case, this would require to know the treatment effect on individual functions with given task difficulty and the exact weights of the individual functions in the composite scales as well as the time-dependency of the individual functions.

For example: a treatment effect of 30% improvement in function f_1 or function f_2 yields quite different effect sizes for early and late patients as assessed with scales F_A or F_B with results between 0.4624 and 0.6039 (Table 1).

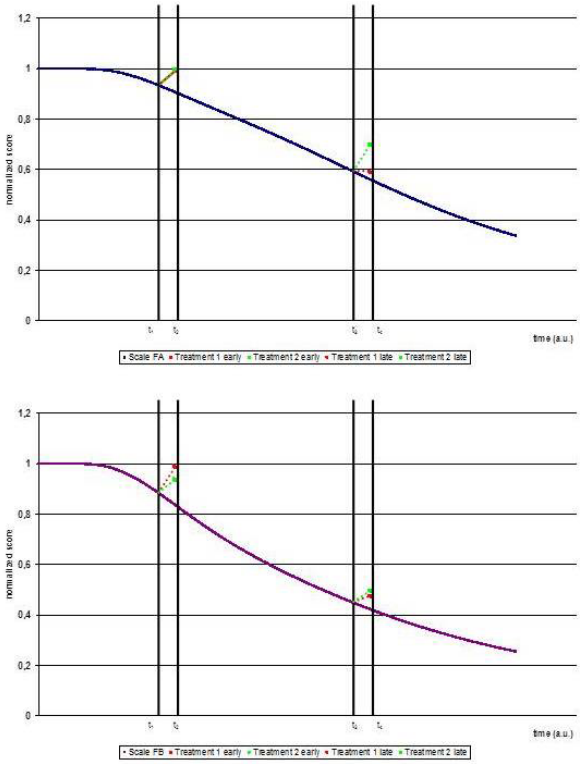

Figure 3 Treatment effects. Functions f_1 and f_2 and Scales F_A and F_B as in Figure 2. A treatment effect of 30% is assumed for f_1 (Treatment 1) or f_2 (Treatment 2). Upper panel) Effects on Scale F_A at early and late time points. Lower panel) Effects on Scale F_B at early and late time points. The graph shows normalized scores over time. The graph shows that the size of the treatment effect depends on the scale that is used.

Inductive mathematical proof

If we assume the average progression of a disease with regard to two instruments within some specified period of time can be described by $f_i(t; a_i, b_i, c_i) = \left(a_i + b_i t e^{c_i/t}\right)^{-1}$, $i = 1, 2$, $t_{min} \le t \le t_{max}$, and that for any time t the underlying distribution of the random variable $X_i(t)$ is a normal one with mean $\mu_i(t) := f_i(t)$.

Table 1 Calculation of effect sizes (Cohen's d) for early and late treatment as assessed with scale F_A and F_B.

	Scale F_A	Scale F_B
Treatment 1 early	0.4796	0.5693
Treatment 1 late	0.5736	0.6039
Treatment 2 early	0.5579	0.4624
Treatment 2 late	0,6005	0.5681

In scale F_A: $F_A(t; a_i, b_i, c_i, i = 1, 2) = 3/8\ f_1(t; a_1, b_1, c_1) + 5/8\ f_2\ (t; a_2, b_2, c_2)$. In scale F_B: $F_B\ (t; a_i, b_i, c_i, i = 1, 2) = 2/3\ f_1(t; a_1, b_1, c_1) + 1/3\ f_2\ (t; a_2, b_2, c_2)$. A treatment effect of 30% is assumed for f_1 (Treatment 1) or f_2 (Treatment 2).

For its standard deviation $\sigma_i(t)$ we assume that always a percentage of 1- α of the distribution has a relative deviation from the mean from at most β percent. To be more precise, if $z_{\alpha/2}$ denotes the (1-α/2) -quantile of the standard normal distribution, then $\sigma_i(t)$ can be determined by the equations $1-\alpha = P\left(\left|\frac{X_i(t)-\mu_i(t)}{\sigma_i(t)}\right| \le z_{\alpha/2}\right)$ $= P\left(|X_i(t)-\mu_i(t)| \le \underbrace{\sigma_i(t)\cdot z_{\alpha/2}}_{=\beta\mu_i(t)}\right)$, hence $\sigma_i(t) = \frac{\beta\mu_i(t)}{z_{\alpha/2}}$, whence for any time t we have $X_i(t) \sim N(\mu_i(t), \sigma_i^2(t)) = N\left(\mu_i(t), \left(\beta\mu_i(t)/z_{\alpha/2}\right)^2\right)$.

While the above models the case of untreated patients the effect of a proper medication is expressed by scaling factors $1 + \delta_i$, $i = 1, 2$, i.e., on the average the progression of the disease for the treated group is described by $(1 + \delta_i)\, f_i(t; a_i, b_i, c_i)$, $i = 1, 2$, $t_{\min} \le t \le t_{\max}$, where we assume like before that for any time t the random variable $X_i^{\delta_i}(t)$ that describes the observed data at time t is again normally distributed with mean $(1 + \delta_i)\, \mu_i(t)$ and, since the calculation of Cohen's d requires unchanged standard deviations, the same standard deviation like before, i.e., $\sigma_i(t) = \frac{\beta\mu_i(t)}{z_{\alpha/2}}$.

Accepting the assumptions made above we obtain the following result for the effect size "Cohen's d" $d_i(t)$ of the treatment at time t for instrument i, $i = 1,2$.

Theorem 1

The effect size Cohen's d is independent of the time of observation, i.e., $d_i(t) \equiv d_i$.

Proof 1

From the definition of Cohen's d we straightforward obtain

$$d_i(t) = \frac{(1+\delta_i)\,\mu_i(t) - \mu_i(t)}{\sigma_i(t)} = \frac{\delta_i\mu_i(t)}{\beta\mu_i(t)/z_{\alpha/2}} \equiv \frac{z_{\alpha/2}\delta_i}{\beta}.$$

Next consider the case that we are interested in the composed function

$$f(t; a_i, b_i, c_i, \lambda_i, i = 1, 2) = \lambda_1 f_1(t; a_1, b_1, c_1) + \lambda_2 f_2(t; a_2, b_2, c_2),$$

where λ_1, λ_2 are non-negative scaling factors with, say, $\lambda_1 + \lambda_2 = 1$. From an intuitive point of view we expect

$$d(t) = \lambda_1 d_1(t) + \lambda_2 d_2(t) \equiv \lambda_1\frac{z_{\alpha/2}\delta_1}{\beta} + \lambda_2\frac{z_{\alpha/2}\delta_2}{\beta} = \frac{z_{\alpha/2}}{\beta}(\lambda_1\delta_1 + \lambda_2\delta_2)$$

for the effect size $d(t)$ of the composed scale. And in case our intuition is correct, time-independence as a desirable prerequisite for meta-analysis on, say, additive scales would immediately follow.

To compute $d(t)$ for $f(t; a_i, b_i, c_i, \lambda_i, i = 1, 2)$ we have to consider the random variable $X(t) = \lambda_1 X_1(t) + \lambda_2 X_2(t)$ for untreated patients and $X^{\delta}(t) = \lambda_1 X_1^{\delta_1}(t) + \lambda_2 X_2^{\delta_2}(t)$ for treated patients. Obviously, both variables are normally distributed with mean $\mu(t) = \lambda_1\mu_1(t) + \lambda_2\mu_2(t)$ and $\mu^{\delta}(t) = \lambda_1(1 + \delta_1)\mu_1(t) + \lambda_2(1 + \delta_2)\mu_2(t)$ respectively. For the variance $\sigma^2(t)$ of $X(t)$ and hence by assumption also of $X^{\delta}(t)$, we have the basic formula

$$\sigma^2(t) = \lambda_1^2\sigma_1^2(t) + 2\lambda_1\lambda_2 cor\left(X_1(t), X_2(t)\right)\sigma_1(t)\sigma_2(t) + \lambda_2^2\sigma_2^2(t),$$

where $cor\,(X_1(t), X_2(t))$ denotes the correlation of $X_1(t)$, and $X_2(t)$,.

In the general case, i.e. without any restrictions on the correlation we obtain time-dependence on the effect size $d(t)$ of the composed scale. To be more precise, we have

$$d(t) = \frac{\mu^{\delta}(t) - \mu(t)}{\sigma(t)} = \frac{\lambda_1\delta_1\mu_1(t) + \lambda_2\delta_2\mu_2(t)}{\sqrt{\lambda_1^2\sigma_1^2(t) + 2\lambda_1\lambda_2 cor\left(X_1(t), X_2(t)\right)\sigma_1(t)\sigma_2(t) + \lambda_2^2\sigma_2^2(t)}}.$$

To become more specific and to answer the question, whether time-independence can be guaranteed also for composed scales under special assumptions we consider as a simple example the case $cor\,(X_1(t), X_2(t)) = 1$ This assumption yields

$$\sigma^2(t) = \lambda_1^2\sigma_1^2(t) + 2\lambda_1\lambda_2\sigma_1(t)\sigma_2(t) + \lambda_2^2\sigma_2^2(t) = \left(\lambda_1\sigma_1(t) + \lambda_2\sigma_2(t)\right)^2,$$

hence $\sigma(t) = (\lambda_1\sigma_1(t) + \lambda_2\sigma_2(t))$ and we can calculate Cohen's d:

$$d(t) = \frac{\mu^{\delta}(t) - \mu(t)}{\sigma(t)} = \frac{\lambda_1(1+\delta_1)\mu_1(t) + \lambda_2(1+\delta_2)\mu_2(t) - \left(\lambda_1\mu_1(t) + \lambda_2\mu_2(t)\right)}{\lambda_1\sigma_1(t) + \lambda_2\sigma_2(t)} = \frac{\lambda_1\delta_1\mu_1(t) + \lambda_2\delta_2\mu_2(t)}{\lambda_1\sigma_1(t) + \lambda_2\sigma_2(t)}.$$

Using $\sigma_i(t) = \frac{\beta\mu_i(t)}{z_{\alpha/2}}$ we finally obtain $d(t) = \frac{z_{\alpha/2}}{\beta}\cdot\frac{\lambda_1\delta_1\mu_1(t) + \lambda_2\delta_2\mu_2(t)}{\lambda_1\mu_1(t) + \lambda_2\mu_2(t)}$, which is in general still not independent of the time t.

In order to further analyze the dependence of the "composed Cohen's d" on the involved parameters we rewrite its formula. Under the assumption on standard deviations and correlation made above we obtain for the effect size:

Theorem 2

$$d(t) = d(t, \lambda_1, \lambda_2, \delta_1, \delta_2) = \frac{z_{\alpha/2}}{\beta}\cdot\left(\delta_1 + (\delta_2 - \delta_1)\frac{1}{1 + \left(\lambda_1\mu_1(t)/\lambda_2\mu_2(t)\right)}\right).$$

Proof 2

We calculate

$$\begin{aligned} d(t) &= \frac{z_{\alpha/2}}{\beta}\cdot\frac{\lambda_1\delta_1\mu_1(t) + \lambda_2\delta_2\mu_2(t)}{\lambda_1\mu_1(t) + \lambda_2\mu_2(t)} = \frac{z_{\alpha/2}}{\beta}\cdot\frac{\lambda_1\delta_1\mu_1(t) + \lambda_2\delta_1\mu_2(t) - \lambda_2\delta_1\mu_2(t) + \lambda_2\delta_2\mu_2(t)}{\lambda_1\mu_1(t) + \lambda_2\mu_2(t)} \\ &= \frac{z_{\alpha/2}}{\beta}\cdot\frac{\delta_1\left(\lambda_1\mu_1(t) + \lambda_2\mu_2(t)\right) + (\delta_2 - \delta_1)\lambda_2\mu_2(t)}{\lambda_1\mu_1(t) + \lambda_2\mu_2(t)} \\ &= \frac{z_{\alpha/2}}{\beta}\cdot\left(\delta_1 + (\delta_2 - \delta_1)\frac{\lambda_2\mu_2(t)}{\lambda_1\mu_1(t) + \lambda_2\mu_2(t)}\right) = \frac{z_{\alpha/2}}{\beta}\cdot\left(\delta_1 + (\delta_2 - \delta_1)\frac{1}{1 + \left(\lambda_1\mu_1(t)/\lambda_2\mu_2(t)\right)}\right) \\ &= d(t, \lambda_1, \lambda_2, \delta_1, \delta_2). \end{aligned}$$

From a theoretical point of view we can now observe the following:

1) If $\delta_1 = \delta_2$, then Cohen's d of the composed measure is independent of the time and in particular equals the weighted sum of the effect sizes d_1 and d_2, i.e.,

$$d(t) \equiv \frac{z_{\alpha/2}}{\beta} \cdot \delta_1 = \frac{z_{\alpha/2}}{\beta} \cdot \delta_1 (\lambda_1 + \lambda_2) = \frac{z_{\alpha/2}}{\beta} (\lambda_1 \delta_1 + \lambda_2 \delta_2).$$

2) If $\lambda_i = 0$, i $\in \{1, 2\}$, then Cohen's d of the composed measure is independent of the time, to be more precise $d(t) \equiv \frac{z_{\alpha/2}}{\beta} \cdot \delta_i$. (Actually this reflects that the choice of parameter implies that the function of interest is no longer a composed one.)

The second observation straightforward leads to the question whether the choices of $\lambda_i = 0$, i $\in \{1, 2\}$ are the extreme ones concerning $d(t)$ over the domain D: = $\{\lambda$: = $(\lambda_1, \lambda_2) | \lambda_1, \lambda_2 \geq 0; \lambda_1 + \lambda_2 = 1\}$?

Theorem 3 $\frac{z_{\alpha/2}}{\beta} \cdot \min\{\delta_1, \delta_2\} \leq \min_{\lambda \in D} d(t, \lambda) \leq \max_{\lambda \in D} d(t, \lambda) \leq \frac{z_{\alpha/2}}{\beta} \cdot \max\{\delta_1, \delta_2\}$.

Proof 3

Without loss of generality assume that $\delta_1 \leq \delta_2$. Then it follows on the one side

$$d(t) = \frac{z_{\alpha/2}}{\beta} \cdot \left(\delta_1 + (\delta_2 - \delta_1) \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) \leq \frac{z_{\alpha/2}}{\beta} \cdot (\delta_1 + (\delta_2 - \delta_1)) = \frac{z_{\alpha/2}}{\beta} \cdot \delta_2 = \frac{z_{\alpha/2}}{\beta} \cdot \max\{\delta_1, \delta_2\}$$

and on the other side

$$d(t) = \frac{z_{\alpha/2}}{\beta} \cdot \left(\delta_1 + (\delta_2 - \delta_1) \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) \geq \frac{z_{\alpha/2}}{\beta} \cdot \delta_1 = \frac{z_{\alpha/2}}{\beta} \cdot \min\{\delta_1, \delta_2\}.$$

Note that we have always equality if $\delta_1 = \delta_2$ which reflects the first observation made above, hence scaling cannot change the effect size. However, if, say, $\delta_1 < \delta_2$, then Cohen's d can be changed by a factor of up to δ_2 / δ_1 by choosing different scales.

Next let us consider the situation that either $\delta_1 = 0$ or $\delta_2 = 0$.

Corollary 1 Under the assumption made above on standard deviations and correlation we obtain for the effect size $d(t) = \frac{d_1}{1 + \left(\lambda_2 \mu_2(t) / \lambda_1 \mu_1(t)\right)}$ if $\delta_1 \neq 0 = \delta_2$ and $d(t) = \frac{d_2}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)}$ if $\delta_1 = 0 \neq \delta_2$.

Proof

First note that $d_i(t) = \frac{z_{\alpha/2} \delta_i}{\beta}$ is equivalent to $\frac{d_i(t)}{\delta_i} = \frac{z_{\alpha/2}}{\beta}$.

Hence, using Theorem 2 we obtain

$$d(t) = \frac{z_{\alpha/2}}{\beta} \cdot \left(\delta_1 + (\delta_2 - \delta_1) \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) = \frac{d_i}{\delta_i} \cdot \left(\delta_1 + (\delta_2 - \delta_1) \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right)$$

for $i \in \{1, 2\}$.

If $\delta_1 = 0 \neq \delta_2$ we conclude

$$d(t) = \frac{z_{\alpha/2}}{\beta} \cdot \left(\delta_1 + (\delta_2 - \delta_1) \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) = \frac{d_2}{\delta_2} \cdot \left(0 + (\delta_2 - 0) \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) = \frac{d_2}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)}.$$

If $\delta_1 \neq 0 = \delta_2$ we conclude

$$d(t) = \frac{d_1}{\delta_1} \cdot \left(\delta_1 - \delta_1 \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) = d_1 \left(1 - \frac{1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) = d_1 \left(\frac{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right) - 1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} \right) = \frac{d_1}{\left(\lambda_2 \mu_2(t) / \lambda_1 \mu_1(t)\right) + 1}.$$

Finally let us compare in the situations $\delta_1 = 0$ or $\delta_2 = 0$ the composed Cohen's d with the intuitive choice $d(t) = \lambda_i d_i$.

Corollary 2 Under the assumption made above on standard deviations and correlation and assuming $\mu_1(t) < \mu_2(t)$ für $t \in \{t_{\min}, t_{\max}\}$ we obtain for the effect size

$$d(t) < \lambda_1 d_1 \text{ if } \delta_1 \neq 0 = \delta_2 \text{ and } d(t) > \lambda_2 d_2 \text{ if } \delta_1 = 0 \neq \delta_2.$$

Proof

Using Corollary 1 for the case $\delta_1 \neq 0 = \delta_2$ we obtain

$$d(t) = \frac{d_1}{1 + \left(\lambda_2 \mu_2(t) / \lambda_1 \mu_1(t)\right)} = \frac{\lambda_1 \mu_1(t)}{\lambda_1 \mu_1(t) + \lambda_2 \mu_2(t)} d_1 < \frac{\lambda_1 \mu_1(t)}{\lambda_1 \mu_1(t) + \lambda_2 \mu_1(t)} d_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} d_1 = \lambda_1 d_1.$$

And in the case $\delta_1 = 0 \neq \delta_2$ we obtain

$$d(t) = \frac{d_1}{1 + \left(\lambda_1 \mu_1(t) / \lambda_2 \mu_2(t)\right)} = \frac{\lambda_2 \mu_2(t)}{\lambda_1 \mu_1(t) + \lambda_2 \mu_2(t)} d_1 > \frac{\lambda_2 \mu_2(t)}{\lambda_1 \mu_2(t) + \lambda_2 \mu_2(t)} d_1 = \frac{\lambda_2}{\lambda_1 + \lambda_2} d_1 = \lambda_2 d_1.$$

Discussion

Rather than drawing conclusions from clinical trials via the differences in the cumulative scores of clinical scales it has become a custom to calculate effect sizes. The intention being to allow comparison of the effect of treatments in the same indication but whilst using different instruments. Using meta-analytic procedures a pooled effect size then is calculated. Meta-analyses are assumed to be the tools to achieve an unbiased analysis of disease severity and the efficacy of treatments [1-4]. Meta-analyses thus are used to summarize results across studies and even across different indications. Considering the multitude of clinical trials and the multitude of treatments such methods are urgently needed and with certain study designs and endpoints this may be an appropriate procedure. It is one limitation of the present study that modulation of effect size calculation by instruments applied and disease stages analyzed applies only to additive scales. These, however, are used frequently in neurodegenerative disease and it is therefore necessary to be aware of the methodological boundary conditions for calculation of effect sizes for additive scales.

Simulation of decline of function in neurodegenerative disease with a non-linear representation of function

demonstrates that calculation of effect sizes for early and late patients is subject to distortion by differences in the vulnerability of brain tissue or task difficulty and scale construction, respectively. Effect sizes are not inert to disease progression and the instruments used to detect it and therefore do not replace experienced clinical assessment of disease impact and treatment effect. Meta-analyses must not pool effect sizes from clinical trials in patients with different severity of disease. Clearly the use of the same scales across the whole disease process is not possible for reasons of differences in task difficulty creating floor and ceiling effects.

It has already been reported that the ADAS-cog and its subscales provide maximum information at moderate levels of cognitive dysfunction [25,26]. Raw score differences toward the lower and higher ends of the scale corresponded to large differences in cognitive dysfunction, whereas raw score differences toward the middle of the scale corresponded to smaller differences [25]. In more severe stages of dementia the ADAScog loses its sensitivity of change so much that the SIB was developed to assess patients who are unable to complete tests such as the ADAS-cog [18]. However, use of different composite scales is not possible since the subscales are not scaled according to task difficulty, are not balanced across different neuropsychological functions, and are weighted differently in different composite scales. A recent post-hoc analysis of published data is in good harmony with the conclusions from the simulation provided here and the mathematical analysis [27]. In that study [27] it was shown that effect size calculation is subject to an interaction of cognitive domain, disease severity, and instruments used for assessment.

In principle, these distortions by disease stage and treatments affecting different functions within a given scale could be measured and mathematical analysis (above and appendix) shows a way to estimate a scaling factor that needed to be introduced. Analysis of current shortcomings then needs to be extended. In the present model we only assume two functions representing two activities, which yields a scaling factor of up to δ_2/δ_1 (cf. above). Clinical scales such as the MMSE or the ADAS-cog are composed of a multitude of functions. When analyzing the ADAScog, for instance, at least four functions need to be considered: memory, orientation, language, and praxis. Therefore, in order to be able to estimate the relative scaling factors would require a very large population.

It has been suggested to call effect sizes of below 0.2 as 'small and above 0.5 as 'medium' [28]. The above analysis demonstrates that the naïve analysis of composite measures may bring about a false categorization of effect size. Effect size calculation of composite endpoints therefore cannot be used as a guideline for the judgment on therapeutic efficacy for neurobiological and statistical reasons. The numerical value of the analysis depends on the choice of the instrument and is subject to distortion by disease progression. Calculation of effect sizes, therefore, can not substitute for clinical assessment. Clinical expertise determines the choice of the instrument - the results therefore need to be interpreted with clinical expertise. Overall, statistical measures and meta-analyses of additive scales obfuscate, rather than clarify, the evidence on therapeutic efficacy in neurodegenerative disease.

In the past, clinical global assessments were the gold standard by which assessment scales were validated. In other words, scales were devised to act as a good proxy for clinical judgment which could be administered by less experienced clinicians. However, these scales clearly have great difficulties when extended over the range and time course of a degenerative disease. What may be a more satisfactory method of measuring change than combining many less than satisfactory study results would be to design a more sensitive way of capturing the clinical assessment. Clinical assessment uses parallel processing and multiple inputs which can account for variations in severity or even input of carepersons. Perhaps devising a more detailed global assessment with maybe 10 - 15 anchor points on a Likert scale that allows clinicians to provide a far more nuanced assessment than the present 7 (often then condensed to 5) point scale. For example it requires much greater evidence and confidence to move from minimal to major improvement than from no change to minimal improvement in most clinicians view and yet they represent similar degrees of improvement on the typical current global assessment scales. This tendency to conservative no change assessments caused by the lack of sensitivity of the scale may be why in the past the clinicians global assessment, whilst being the standard by which all patients in the real world and all other scales are assessed has not been regarded as a useful tool in clinical trials.

Conclusions

In the face of the clear lack of credibility in pooling effect size calculations on grouped and yet disparate studies for meta-analysis it may be time to put the clinical appraisal that has served for generations back where it belongs as cornerstone of our efficacy assessments and decision making about the utility of treatments in neurodegenerative diseases.

Acknowledgements

The research was performed without external funding.

Author details

[1]Department of Psychiatry and Psychotherapy II, Mental Health & Old Age Psychiatry, Ulm University, Ulm, Germany. [2]Department of Old Age Psychiatry, Southampton University, Southampton, UK. [3]Department of

Psychiatry and Psychotherapy, Technische Universität München, München, Germany. [4]Department of Wirtschafts- und Organisationswissenschaften, Universität der Bundeswehr München, Neubiberg, Germany.

Authors' contributions

MWR, DW, and HF raised the ideas and elaborated the medical content. AB performed the mathematical proof. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 22 February 2011 Accepted: 16 December 2011
Published: 16 December 2011

References

1. Chalmers TC, Sacks H: **Randomized clinical trials in surgery.** *N Engl J Med* 1979, **301**:1182.
2. Chalmers TC: **Meta-analysis in clinical medicine.** *Trans Am Clin Climatol Assoc* 1988, **99**:144-150.
3. Lau J, Chalmers TC: **The rational use of therapeutic drugs in the 21st century. Important lessons from cumulative meta-analyses of randomized control trials.** *Int J Technol Assess Health Care* 1995, **11**:509-522.
4. Sacks HS, Berrier J, Reitman D, ncona-Berk VA, Chalmers TC: **Meta-analyses of randomized controlled trials.** *N Engl J Med* 1987, **316**:450-455.
5. Field AP, Gillett R: **How to do a meta-analysis.** *Br J Math Stat Psychol* 2010, **63**:665-694.
6. Hyman BT, Van Horsen GW, Damasio AR, Barnes CL: **Alzheimer's disease: cell-specific pathology isolates the hippocampal formation.** *Science* 1984, **225**:1168-1170.
7. Hodges JR: **Memory in the dementias.** In *The Oxford Handbook of Memory.* Edited by: Tulving E, Craik FIM. Oxford, New York: Oxford University Press; 2000:441-459.
8. Marshall JC, Fink GR: **Spatial cognition: where we were and where we are.** *Neuroimage* 2001, **14**:S2-S7.
9. Save E, Poucet B: **Hippocampal-parietal cortical interactions in spatial cognition.** *Hippocampus* 2000, **10**:491-499.
10. Godefroy O, Cabaret M, Petit-Chenal V, Pruvo JP, Rousseaux M: **Control functions of the frontal lobes. Modularity of the central-supervisory system?** *Cortex* 1999, **35**:1-20.
11. Nagahama Y, Okada T, Katsumi Y, Hayashi T, Yamauchi H, Oyanagi C, *et al*: **Dissociable mechanisms of attentional control within the human prefrontal cortex.** *Cereb Cortex* 2001, **11**:85-92.
12. Rowe JB, Toni I, Josephs O, Frackowiak RS, Passingham RE: **The prefrontal cortex: response selection or maintenance within working memory?** *Science* 2000, **288**:1656-1660.
13. Gron G, Bittner D, Schmitz B, Wunderlich AP, Riepe MW: **Subjective memory complaints: objective neural markers in patients with Alzheimer's disease and major depressive disorder.** *Ann Neurol* 2002, **51**:491-498.
14. Gron G, Riepe MW: **Neural basis for the cognitive continuum in episodic memory from health to Alzheimer disease.** *Am J Geriatr Psychiatry* 2004, **12**:648-652.
15. Bittner D, Gron G, Schirrmeister H, Reske SN, Riepe MW: **[18F]FDG-PET in patients with Alzheimer's disease: marker of disease spread.** *Dement Geriatr Cogn Disord* 2005, **19**:24-30.
16. Rosen WG, Mohs RC, Davis KL: **A new rating scale for Alzheimer's disease.** *Am J Psychiatry* 1984, **141**:1356-1364.
17. Folstein MF, Folstein SE, McHugh PR: **"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.** *J Psychiatr Res* 1975, **12**:189-198.
18. Saxton J, Swihart AA: **Neuropsychological assessment of the severely impaired elderly patient.** *Clin Geriatr Med* 1989, **5**:531-543.
19. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J: **The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia.** *Neurology* 1994, **44**:2308-2314.
20. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW: **Studies of illness in the aged. The index of adl: A standardized measure of biological and psychosocial function.** *JAMA* 1963, **185**:914-919.
21. Ulrich M, Jonas C, Gron G: **Functional compensation of increasing memory encoding demands in the hippocampus.** *Neuroreport* 2010, **21**:59-63.
22. Mendiondo MS, Ashford JW, Kryscio RJ, Schmitt FA: **Modelling mini mental state examination changes in Alzheimer's disease.** *Stat Med* 2000, **19**:1607-1616.
23. Ashford JW, Kolm P, Colliver JA, Bekian C, Hsu LN: **Alzheimer patient evaluation and the mini-mental state: item characteristic curve analysis.** *J Gerontol* 1989, **44**:139-146.
24. Ashford JW, Shan M, Butler S, Rajasekar A, Schmitt FA: **Temporal quantification of Alzheimer's disease severity: 'time index' model.** *Dementia* 1995, **6**:269-280.
25. Benge JF, Balsis S, Geraci L, Massman PJ, Doody RS: **How well do the ADAS-cog and its subscales measure cognitive dysfunction in Alzheimer's disease?** *Dement Geriatr Cogn Disord* 2009, **28**:63-69.
26. Panisset M, Roudier M, Saxton J, Boller F: **Severe impairment battery. A neuropsychological test for severely demented patients.** *Arch Neurol* 1994, **51**:41-45.
27. Riepe MW, Janetzky W, Lemming OM: **Measuring therapeutic efficacy in patients with Alzheimer's disease: role of instruments.** *Dement Geriatr Cogn Disord* 2011, **31**:233-238.
28. Cohen J: *Statistical power analysis for the behavioral sciences.* 2 edition. Lawrence Erlbaum Associates; 1988.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2288/11/169/prepub

doi:10.1186/1471-2288-11-169
Cite this article as: Riepe *et al.*: Additive scales in degenerative disease - calculation of effect sizes and clinical judgment. *BMC Medical Research Methodology* 2011 **11**:169.