A limitation of single-subject designs is the difficulty of generalizing their results to other subjects, because of the small number of subjects that are investigated. To enhance generalizability, researchers replicate across subjects, either within studies, such as in a multiple-baseline design, or across studies. As more replications emerge and the evidence base accumulates so does the need for statistical methods designed to synthesize single-subject experimental design studies’ results and to explore sources of systematic variability in single-subject results by employing moderator analyses. One approach for combining single-subject data within and across studies is multilevel modeling. If we have several single-subject studies, with at least in some studies more than one subject, three hierarchical levels can be distinguished: measurements are grouped in subjects and subjects are in turn grouped in studies. The hierarchical structure is illustrated in Figure 1.

Figure 1. The three-level hierarchical structure for single-subject experimental design studies.

A multilevel model that can be used to combine single-subject data is an extension of the regression model of Center, Skiba and Casey (1985-1986). More specifically, Van den Noortgate and Onghena (2003a, 2003b, 2008) suggest to use a hierarchical model in which individual measurements are regressed on a time indicator, $T$, a dummy variable for the treatment phase, $D$, and an interaction term of these variables, $DT$:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}T_{ijk} + \beta_{2jk}D_{ijk} + \beta_{3jk}T'_{ijk}D_{ijk} + e_{ijk} \quad \text{with} \quad e_{ijk} \sim N(0, \sigma^2)$$

and $i$ standing for the measurement occasion ($i = 0, 1, \ldots, I$), $j$ for the subject ($j = 0, 1, \ldots, J$) and $k$ for the study ($k = 0, 1, \ldots, K$).

Equation 1 shows that in the baseline phase the expected score for the $j$th subject in study $k$, this is $\hat{Y}_{ijk}$, equals $\beta_{0jk} + \beta_{1jk}T_{ijk}$, while it is $(\beta_{0jk} + \beta_{2jk}) + (\beta_{1jk} + \beta_{3jk})T_{ijk}$ in the treatment phase (see Figure 2).
Figure 2. Regression model to analyze data from single-subject AB phase design.

$\hat{y}_{ijk} = \beta_{0jk} + \beta_{1jk} T_{ijk}$  
$\hat{y}_{ijk} = (\beta_{0jk} + \beta_{2jk}) + (\beta_{1jk} + \beta_{3jk}) T_{ijk}$

$\beta_{0jk}$ therefore indicates the expected baseline level at the start of the baseline phase (when $T = 0$), and $\beta_{1jk}$ the linear time trend in the baseline scores. If the time indicator in the interaction term, $D_{ijk} T_{ijk}'$, is centered around the first observation of the treatment phase, the coefficient $\beta_{2jk}$ can then be interpreted as the immediate effect of the treatment on the outcome, whereas $\beta_{3jk}$ gives an indication of the effect of the treatment on the time trend.

At the second level of the model, the variation over subjects is described using four equations:

$$
\begin{cases}
\beta_{0jk} = \theta_{00k} + u_{0jk} \\
\beta_{1jk} = \theta_{10k} + u_{1jk} \\
\beta_{2jk} = \theta_{20k} + u_{2jk} \\
\beta_{3jk} = \theta_{30k} + u_{3jk}
\end{cases}
$$

with

$$
\begin{bmatrix}
\beta_{0jk} \\
\beta_{1jk} \\
\beta_{2jk} \\
\beta_{3jk}
\end{bmatrix} =
\begin{bmatrix}
\theta_{00k} \\
\theta_{10k} \\
\theta_{20k} \\
\theta_{30k}
\end{bmatrix}
+ 
\begin{bmatrix}
u_{0jk} \\
u_{1jk} \\
u_{2jk} \\
u_{3jk}
\end{bmatrix} 
\sim N(0, \Sigma_u),
$$

The first equation indicates that the baseline performance for subject $j$ from study $k$ equals an overall baseline performance for study $k$, plus a random deviation from this mean; the subsequent equations describe the variation over subjects from the same study of the time effect in the baseline condition, the immediate treatment effect, and the treatment effect on the linear trend, respectively.

At the third level, the variation of the study-specific regression coefficients from the second level equations is described:
\[
\begin{align*}
\theta_{00k} &= y_{000} + v_{00k} \\
\theta_{10k} &= y_{100} + v_{10k} \\
\theta_{20k} &= y_{200} + v_{20k} \\
\theta_{30k} &= y_{300} + v_{30k}
\end{align*}
\]

with \[
\begin{bmatrix}
v_{00k} \\
v_{10k} \\
v_{20k} \\
v_{30k}
\end{bmatrix} \sim N(0, \Sigma_v)
\] (3)

Residuals at all three levels are assumed to be multivariate normally distributed. Parameters of interest are typically primarily the \(\gamma\)'s, in multilevel literature called fixed effects, in this case referring to the mean regression coefficients, as well as the (co)variation in the regression coefficients over subjects or studies, in multilevel literature called the variance components.

Moeyaert, Ugille, Ferron, Beretvas and Van den Noortgate (2012) demonstrated that this approach results in unbiased estimates of the immediate treatment effect and the treatment effect on the time trend.

References:


