suddengains: An R package to identify sudden gains in longitudinal data

Milan Wiedemann[1,2], Graham R Thew[1,2,3], Richard Stott[4], & Anke Ehlers[1,2,4]

[1] Department of Experimental Psychology, University of Oxford, Oxford, UK

[2] Oxford Health NHS Foundation Trust, Oxford, UK

[3] Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[4] King's College London, London, UK

## Author Note

Correspondence concerning this article should be addressed to Milan Wiedemann, Department of Experimental Psychology, University of Oxford. E-mail: milan.wiedemann@gmail.com

## Abstract

Sudden gains are large and stable changes on an outcome variable between consecutive measurements, for example during a psychological intervention with multiple assessments. Researching these occurrences could help understand individual change processes in longitudinal data. Tang and DeRubeis (1999) suggested three criteria to define sudden gains in psychological interventions. However, identifying sudden gains based on these criteria can be time consuming and prone to errors if not fully automated. Further, methodological decisions such as how missing data, or multiple gains, are handled vary across studies and are reported with different levels of detail. These problems limit the comparability of individual studies and make it hard to understand or replicate the exact methods used. The R package *suddengains* provides a set of tools to facilitate sudden gains studies by automating their identification, adjusting for missing data, and providing descriptive statistics as well as exportable data files for further analysis.

*Keywords:* sudden gains, sudden losses, R, mechanisms of change, processes of change, psychological therapy, longitudinal analysis, single case designs

suddengains: An R package to identify sudden gains in longitudinal data

## Introduction

A sudden gain is a large and stable change in an outcome variable experienced by an individual participant between two consecutive measurement points of a longitudinal data series. They were first defined and investigated by Tang and DeRubeis (1999), who examined session by session changes in depression symptoms among participants undertaking cognitive behavioural therapy. The majority of sudden gains studies to date have been in relation to psychological therapies for depression and anxiety (see Aderka, Nickerson, Boe, & Hofmann, 2012), but the analytic approach could be applied to any longitudinal dataset with regular repeated measurement, usually monitoring the impact of an intervention. In a meta-analysis of 16 studies (total $N = 1104$), Aderka et al. (2012) found that experiencing a sudden gain was associated with better overall clinical outcomes at the end of treatment and at follow-up compared to those who did not experience gains. Given this potential clinical significance, the study of sudden gains may be informative in understanding when and why people experience such large improvements, which could help to increase the efficacy and efficiency of the intervention.

## Identification of sudden gains

Following Tang and DeRubeis (1999), the three criteria used to identify sudden gains are (see also Tang, DeRubeis, Beberman, & Pham, 2005):

1. The gain must be large in absolute terms. While this was originally operationalised as a decrease of at least 7 points on the Beck Depression Inventory (BDI; Beck & Steer, 1993), subsequent studies have used the reliable change index (RCI; Jacobson & Truax, 1991) of the scale to define an appropriate cut-off (see Stiles et al., 2003), following the

formulae below[1] :

$$\text{RCI} = \frac{\text{pre - post}}{S_{\text{diff}}}$$

$$S_{\text{diff}} = \sqrt{2 \times (S_E)^2}$$

$$S_E = s_1\sqrt{1 - r_{xx}} \;\; \text{adapted to:} \;\; S_E = SD_{\textbf{pre}}\sqrt{1 - \alpha_{pre}}$$

$$\text{Cut-off value} = \text{RCI} \times 1.96$$

2. The gain must be large in relative terms. This is defined as a drop of at least 25% of the previous score.

3. The gain must be large relative to symptom fluctuation. The difference between the mean scores of three measurements before the gain, and the three measurements after the gain, must be greater than the pooled standard deviation of these two groups times 2.776 (i.e. the two-tailed critical $t$ value for $\alpha = 0.05$ and $df = 4$).

**Missing data**

Missing data around the period of a potential sudden gain need to be considered carefully, as depending on the number and distribution of missing datapoints for an individual, it may not be possible to calculate sudden gains. Specifically, in order to estimate the standard deviation values in criterion 3, at least two of the three measurements immediately prior to the gain must be present, as well as at least two of the three measurements immediately following the gain. The critical $t$ values used in criterion 3 can be adjusted for missingness as follows (e.g., Lutz et al., 2013): Where no data are missing

---

[1] The RCI formula was modified following Stiles et al. (2003) so that all statistics could be extracted from the sample data. The standard deviation of the normal population ($s_1$) is substituted with the standard deviation of the clinical sample at baseline ($SD_{pre}$), and the test-retest reliability ($r_{xx}$) with the internal consistency at baseline (Cronbach's $\alpha_{pre}$).

$t_{(4;97.5\%)} > 2.776$; where one datapoint is missing either before or after the gain $t_{(3;97.5\%)} >$ 3.182; and where one datapoint is missing both before and after the gain $t_{(2;97.5\%)} > 4.303$.

## Terminology

The naming of specific sessions (or measurement points) around the gain follows the convention that the session immediately prior to the gain is session $N$ (also known as the *pregain* session), and the session immediately after is session *N+1* (or *postgain* session). Other sessions are referred to in relation to session $N$ (e.g. *N-2, N+3*).

## Reversals

According to Tang and DeRubeis (1999) a sudden gain is counted as reversed if 50% of the improvement made during the gain was lost at any subsequent point. For example, where the sudden gain represents a drop in score from 40 to 30 points, the gain is classed as having reversed if a score of 35 or more is observed at any later session.

## Sudden losses

Although less frequently studied than sudden gains, sudden losses represent the inverse phenomenon, where a participant shows a large and stable increase of scores on the outcome variable. While some authors use the inverse criteria of the sudden gains (e.g., Krüger et al., 2014; Lutz et al., 2013), others adjust the percentage threshold of the second criterion (e.g. 33% in König, Karl, Rosner, & Butollo, 2014).

## Why is a package needed?

As indicated by the criteria above, identifying sudden gains requires the application of each of the three criteria to each session to session interval, and that this is performed for each individual in a given dataset. A large number of calculations and extensive manipulation of data is therefore involved, particularly in larger datasets. Doing these data manipulations manually (e.g. in Excel) can be extremely time consuming and lead to errors. It also means that certain methodological decisions, such as the handling of missing data, or of participants with multiple gains, may not be addressed sufficiently or in a consistent way across studies. It is hoped that the use of the *suddengains* package will provide faster and more accurate calculations, as well as offering a transparent and consistent method to address these methodological considerations.

## Functions of the *suddengains* package

The *suddengains* package provides a set of functions to calculate the presence of sudden gains (and sudden losses) within a longitudinal dataset, and to provide basic plots and descriptive statistics of the gains. It can also extract scores on secondary outcome or process measures around the period of each gain. Output files (in SPSS, Excel, or CSV formats) arranged by individual gain, or by person can be generated for further analyses in other programs. Table 1 lists and describes the main functions.

Table 1

*Main functions of the suddengains R package.*

| Function | Description |
| --- | --- |
| **Identify sudden gains** | |
| define_crit1_cutoff() | Uses modified RCI formula to determine a cut-off value for criterion 1 |
| identify_sg(), identify_sl() | Identifies sudden gains/losses |
| **Create datasets** | |
| create_bysg(), create_byperson() | Creates a dataset with one row for each sudden gain/loss or one row for each person |
| extract_values() | Extracts values on a secondary measure around the sudden gain/loss |
| **Describe sudden gains** | |
| describe_sg() | Generates summary descriptive statistics |
| plot_sg(), plot_sg_trajectories() | Creates plots of the average sudden gain, or plots of individual case trajectories |
| **Additional functions** | |
| select_cases() | Selects cases to be included in the sudden gains analysis based on different criteria |
| write_bysg(), write_byperson() | Exports CSV, SPSS, Excel, or STATA files of the sudden gains data sets |

*Note.* More details of each function can be found in the package documentation or using the help() function in R.

## Worked example

This demonstration uses an example dataset[2] that was created to illustrate the functions of this package. The data show self-report weekly questionnaire scores for 43 participants who have received psychological therapy for depression. The intervention lasted for 12 sessions, and each participant completed a set of outcome measures at the beginning of each session, including the BDI (Beck & Steer, 1993) and a fictional secondary measure assessing rumination (RQ).

### Preparation of data

The data to be analysed for sudden gains are arranged in wide format i.e. one row per participant, and one column for each questionnaire score at each measurement point. A unique identifier variable also needs to be included. The optional **select_cases()** function can be used to determine cases that may need to be excluded, for example where very few sessions were attended, or where the pattern of missing data is such that it would not be possible to identify any sudden gains (see Table A1).

### Identification of sudden gains

The **identify_sg()** function applies the sudden gains criteria as specified by the user to each session to session interval in the dataset. As shown below, the user specifies: `data`, the dataset to use in wide format; `sg_crit1_cutoff`, the cutoff value to use for criterion 1 (which can be entered manually or calculated using the **define_crit1_cutoff()** function); `sg_crit2_pct`, the percentage change value to use for criterion 2 (25% by default); `sg_crit3`, whether or not to apply the third criterion (`TRUE` by default); `id_var_name`, the

---

[2] The name of the example dataset is `sgdata`. The data will be loaded together with the R package using `library(suddengains)`.

name of the unique identifier variable within the dataset; and **sg_var_list**, a list of the variables representing the span of sessions to be analysed, which is sessions 1 to 12 in this example. Additional options to customise this analysis are discussed in the package documentation. An alternative function, **identify_sl()**, is identical to **identify_sg()** but applies the criteria in the inverse direction to calculate "sudden losses".

```r
# First, load the suddengains R package
# See author notes for a link to package installation instructions
library(suddengains)
# Identify sudden gains in the dataset 'sgdata':
identify_sg(data = sgdata,
            sg_crit1_cutoff = 7,
            sg_crit2_pct = 0.25,
            sg_crit3 = TRUE,
            id_var_name = "id",
            sg_var_list = c("bdi_s1", "bdi_s2", "bdi_s3", "bdi_s4",
                            "bdi_s5", "bdi_s6", "bdi_s7", "bdi_s8",
                            "bdi_s9", "bdi_s10", "bdi_s11", "bdi_s12"))
```

The output shows each between-session interval, for example **sg_2to3**, with sudden gains indicated by a value of 1. To permit further analysis of our data, we wish to obtain an output dataset containing both the original data and the newly identified sudden gains. As participants may experience more than one gain, as in the present example, and to allow for different subsequent analyses, the package provides two options for output datasets: The **create_bysg()** function creates a dataset structured with one row per sudden gain, and the **create_byperson()** function creates a dataset structured with one row per person, indicating whether or not they experienced a sudden gain. The **create_bysg()** function is shown below. The **tx_start_var_name** and **tx_end_var_name** arguments are used to specify

the start and end of treatment (tx) variables, and `sg_measure_name` specifies the name of the measure used to calculate sudden gains. To continue working in another program (e.g. SPSS, STATA, Excel) the functions **write_bysg()** and **write_byperson()** can be used to export the datasets created in R (R Core Team, 2018) as *.sav*, *.dta*, *.xlsx*, or *.csv* files.

```r
# Create output dataset with one row per sudden gain
# and save as an object called "bysg" to use later
bysg <- create_bysg(data = sgdata,
                    sg_crit1_cutoff = 7,
                    sg_crit2_pct = 0.25,
                    sg_crit3 = TRUE,
                    id_var_name = "id",
                    tx_start_var_name = "bdi_s1",
                    tx_end_var_name = "bdi_s12",
                    sg_var_list = c("bdi_s1", "bdi_s2", "bdi_s3",
                                    "bdi_s4", "bdi_s5", "bdi_s6",
                                    "bdi_s7", "bdi_s8", "bdi_s9",
                                    "bdi_s10", "bdi_s11", "bdi_s12"),
                    sg_measure_name = "bdi")
```

**Analysis of sudden gains**

In this example, we have calculated sudden gains based on depression scores using the BDI. In analysing these gains, we are interested in how rumination scores on the fictional RQ measure change around the period of the sudden gains in depression. The **extract_values()** function extracts the RQ values from the three sessions before (*N-2*,

*N-1, N*) and the three sessions after (*N+1, N+2, N+3*) each depression sudden gain.[3] This function can be applied to either the `bysg` or `byperson` dataset. By default the extracted values will be added as new variables to the dataset used. Here we demonstrate applying this function to the `bysg` dataset, as shown in the code below. First, the RQ variables are added to the `bysg` dataset. Second, the **extract_values()** function is applied. Note that the list of RQ variables included in the `extract_var_list` argument must match those used for the `sg_var_list` argument used previously in the **create_bysg()** function. This means that the number of variables in these lists has to be identical and measured at the same time points. The output can be saved as a new object, or the existing `bysg` object can be overwritten, as in this example. The RQ scores now in the `bysg` dataset can be examined, for example to look at the temporal relationship between changes in rumination and changes in depression symptoms.

```r
# 1. Select the ID and variables from a second measure
sgdata_rq <- dplyr::select(sgdata,

                           "id",

                           "rq_s1", "rq_s2", "rq_s3",

                           "rq_s4", "rq_s5", "rq_s6",

                           "rq_s7", "rq_s8", "rq_s9",

                           "rq_s10", "rq_s11", "rq_s12")


# 2. Add the variables in 'sgdata_rq' to the 'bysg' dataset created earlier
bysg <- dplyr::left_join(bysg, sgdata_rq, by = "id")


# 3. Extract values on the second measure around the sudden gain
bysg <- extract_values(data = bysg,
```

---

[3] Note that in code, we refer to these sessions as sg_bdi_**2n**, sg_bdi_**1n**, sg_bdi_**n**, sg_bdi_**n1**, sg_bdi_**n2**, and sg_bdi_**n3**, respectively.

```r
                        id_var_name = "id_sg",

                        extract_var_list = c("rq_s1", "rq_s2", "rq_s3",

                                              "rq_s4", "rq_s5", "rq_s6",

                                              "rq_s7", "rq_s8", "rq_s9",

                                              "rq_s10", "rq_s11", "rq_s12"),

                        extract_measure_name = "rq",

                        add_to_data = TRUE)
```

The **describe_sg()** function provides descriptive statistics about the sudden gains based on the variables from the `bysg` and `byperson` datasets. For the present example, this function indicates that 16 of the 43 participants experienced a sudden gain, and 9 experienced more than one gain, leading to a total of 26 sudden gains within the data. Information on the mean gain magnitude and reversals is also provided.

The **plot_sg()** function plots the "average" sudden gain, and can be used to show the primary or secondary outcome measure data (Figure 1). The `sg_pre_post_var_list` argument specifies the pregain and postgain variables to be plotted, namely sessions *N-2* to *N+3*. This function is built using the R-package *ggplot2* (Version 3.1.1; Wickham, 2016) and additional *ggplot2* functions can be added to the plot. It is also possible to plot the average gain magnitude of different groups (e.g. two treatment arms in a trial) in one figure by using the optional **group** argument (see Figure A1). An additional function, **plot_sg_trajectories()**, is available to plot the trajectories of individuals within the dataset (see Figure A2). This function can be paired with a filter command (e.g. **filter()** from R-package *dplyr* (Version 0.8.0.1; Wickham, François, Henry, & Müller, 2018)) to visualise trajectories of specific groups of participants. For example, all participants with more than one sudden gain, or all participants with a sudden gain between sessions 3 and 4 (see Figure A3).

```
# Create average sudden gain plot for BDI data (see Figure 1a):

plot_sg(data = bysg,

        tx_start_var_name = "bdi_s1",

        tx_end_var_name = "bdi_s12",

        sg_pre_post_var_list = c("sg_bdi_2n", "sg_bdi_1n", "sg_bdi_n",

                                 "sg_bdi_n1", "sg_bdi_n2", "sg_bdi_n3"),

        ylab = "BDI")
```
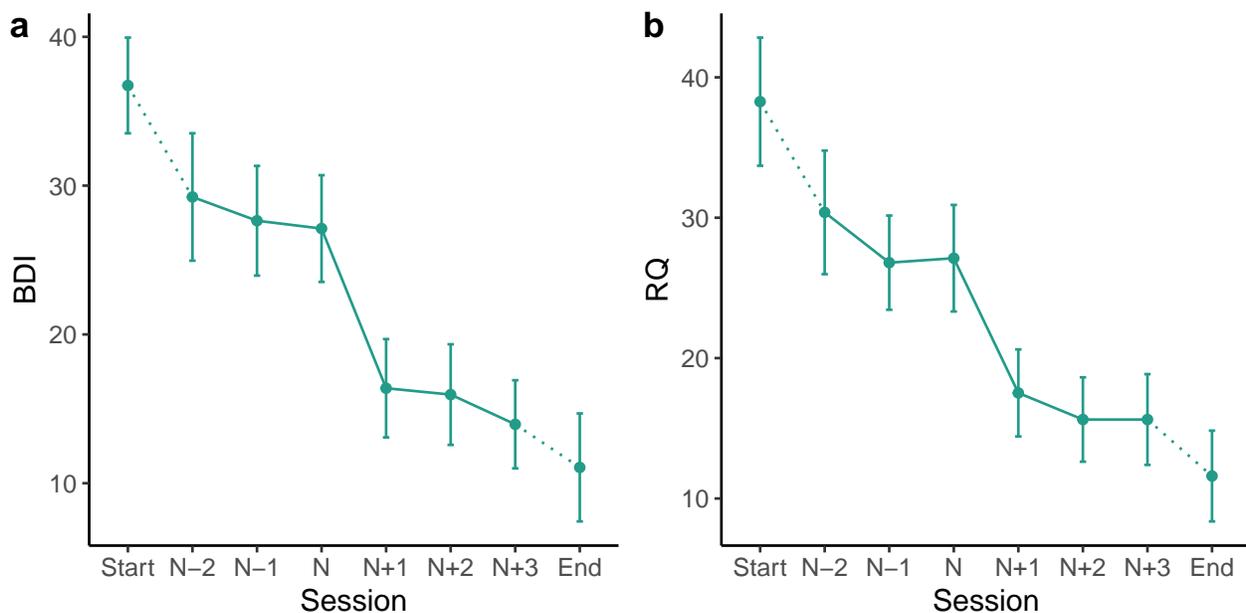


*Figure 1*. (a) Average gain magnitute on the BDI and (b) average change in rumination (RQ) around the time of the sudden gain on BDI for all sudden gains.

## Summary

The analysis of sudden gains provides a detailed examination of individuals' trajectories of change during the course of an intervention, and may help to understand processes of change. The *suddengains* package aims to facilitate the computation of gains, which can be laborious and error-prone. It also aims to address methodological issues such as adjusting for missing data and handling participants with multiple gains. Further

technical details of the package can be found at

https://github.com/milanwiedemann/suddengains, and in the package documentation. The

present examples refer to package version 0.2.1. Readers are also welcome to contact the

authors via email or GitHub at https://github.com/milanwiedemann/suddengains/issues.

# References

Aderka, I. M., Nickerson, A., Boe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93–101. doi:10.1037/a0026455

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Beck, A. T., & Steer, R. A. (1993). *Beck Depression Inventory Manual.* San Antonio, TX: The Psychological Corporation.

Jacobson, N. S., & Truax, P. A. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19. doi:10.1037/0022-006X.59.1.12

König, J., Karl, R., Rosner, R., & Butollo, W. (2014). Sudden gains in two psychotherapies for posttraumatic stress disorder. *Behaviour Research and Therapy, 60*, 15–22. doi:10.1016/j.brat.2014.06.005

Krüger, A., Ehring, T., Priebe, K., Dyer, A. S., Steil, R., & Bohus, M. (2014). Sudden losses and sudden gains during a DBT-PTSD treatment for posttraumatic stress disorder following childhood sexual abuse. *European Journal of Psychotraumatology, 5*, 24470. doi:10.3402/ejpt.v5.24470

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Rottger, M.-A., Jorasz, C., . . . Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23*(1), 14–24. doi:10.1080/10503307.2012.693837

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna,

Austria: R Foundation for Statistical Computing. Retrieved from
https://www.R-project.org/

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., . . . Hardy, G.
E. (2003). Early sudden gains in psychotherapy under routine clinic conditions:
Practice-based evidence. *Journal of Consulting and Clinical Psychology*, *71*(1), 14–21.
doi:10.1037/0022-006X.71.1.14

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in
cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical
Psychology*, *67*(6), 894–904. doi:10.1037/0022-006X.67.6.894

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical
sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of
Consulting and Clinical Psychology*, *73*(1), 168–172. doi:10.1037/0022-006X.73.1.168

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
Retrieved from http://ggplot2.org

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A grammar of data
manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Appendix

Table A1

*Minimum available data patterns required to identify sudden gains.*

| | $x_i$ | $x_{i+1}$ | $x_{i+2}$ | $x_{i+3}$ | $x_{i+4}$ | $x_{i+5}$ |
|---|---|---|---|---|---|---|
| Pattern 1 | ● | ●$_n$ | ● | ● | ○ | ○ |
| Pattern 2 | ● | ●$_n$ | ● | ○ | ● | ○ |
| Pattern 3 | ● | ○ | ●$_n$ | ● | ● | ○ |
| Pattern 4 | ● | ○ | ●$_n$ | ● | ○ | ● |

*Note.* $x_i$ to $x_{i+5}$ represent the scores on the measure used to identify sudden gains at time point $i$ and the subsequent five measurement points. ● represents available data; ●$_n$ represents available data to be considered as the possible pregain session; ○ represents missing data.
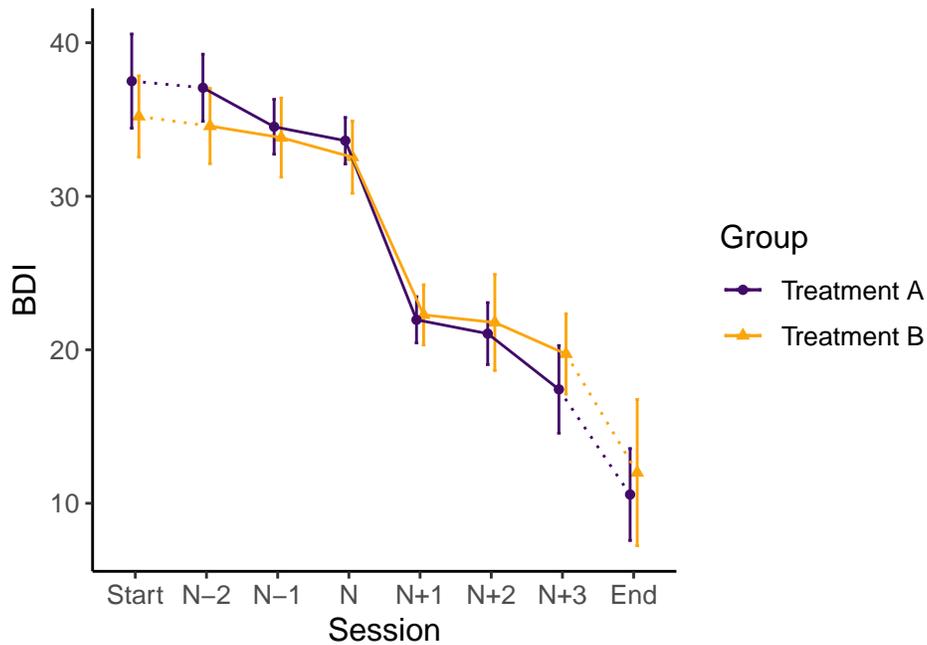


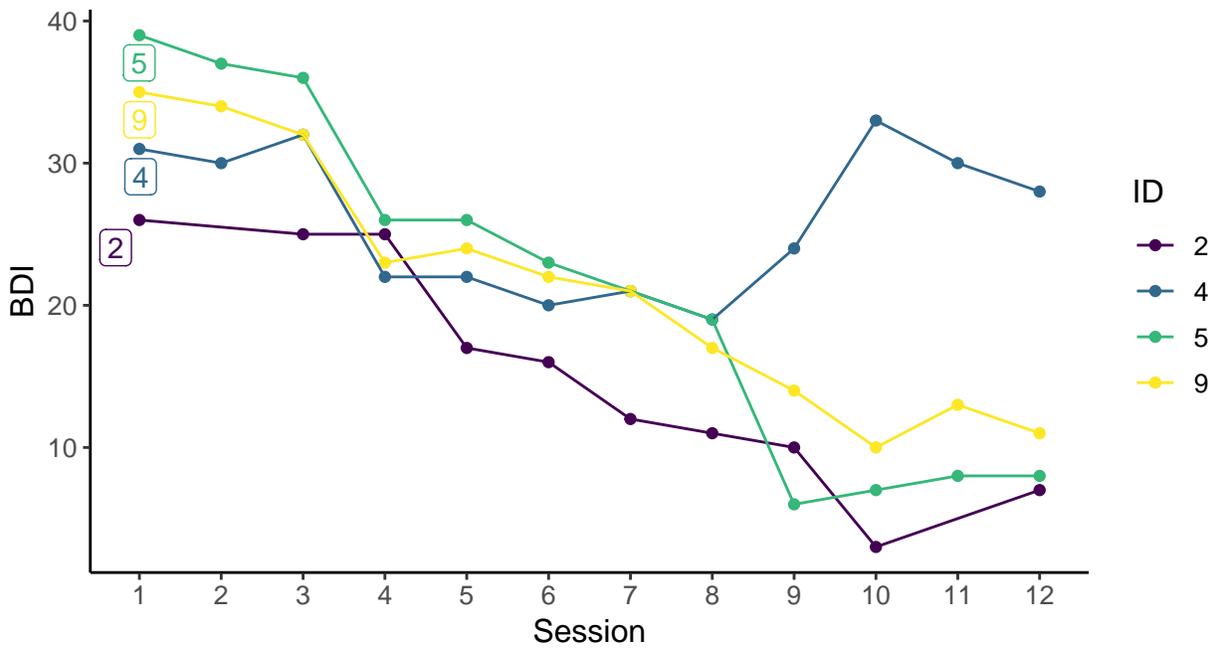*Figure A1*. Average gain magnitude on the BDI for two different treatments.

*Figure A2.* Trajectories of BDI scores for a selection of individual cases.
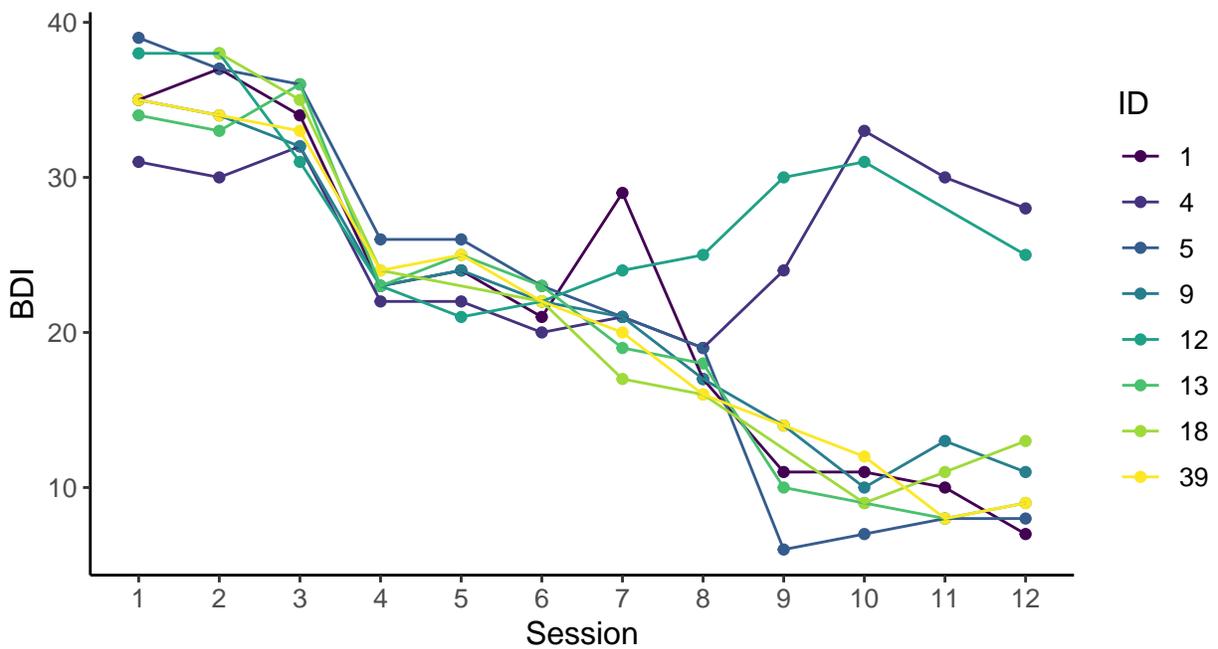


*Figure A3.* Trajectories of BDI scores for all participants with a sudden gain between sessions 3 and 4.