**Supplementary material**

*Separate models or a single model for the three suicidality outcomes?*

First, we tested whether the different outcomes of (i) lifetime suicidal ideation without a lifetime suicide attempt, (ii) lifetime suicidal ideation and/or suicide attempt, and (iii) lifetime suicide attempt show proportional risks (a continuum) with respect to risk factors or not; if not, we modelled them as separate binary outcomes.

We tested whether the effect of TADS score on lifetime suicidal behavior is mediated by MSI score, using the "mediation" package version 4.4.5 for R software version 3.2.2 (2015-08-14) (1-3). This general approach to mediation analysis allows the use of proportional odds logistic regression and logistic regression models for the lifetime suicidal behavior outcome (2, 4). We first tested the proportional odds assumption using R package "ordinal" version 2015.6-28, the null hypothesis being that proportional odds regression model regressing suicidal behavior on TADS score, age and sex does not differ from the model where TADS score is allowed to have nominal effects on lifetime suicidal behavior (i.e., ordering between outcome classes is not constrained). The response/outcome variable was predicted by age, sex, TADS score, BDI score, and MSI score. Our response variable did not satisfy the proportional odds assumption with respect to TADS score (LR statistic = 5.77, df = 1, p = 0.0163), and therefore we regressed the suicidal behavior outcomes separately onto the other covariates using logistic regression models.

*Testing differences in mediation effects*

Possible differences in mediation statistics between the models for the outcomes (i) and (iii) were tested using a random permutation tests with 1000 random permutations [studying moderated mediation was not a feasible strategy because of the "perfect separation" it induced (4, 5)]. The random permutation test is a Monte Carlo test that approximates the null distribution by repeated permutation samples that randomly permute the "ideation only" and "suicide attempt" statuses while holding other characteristics fixed; that is, the procedure tests the hypothesis that these outcomes are "exchangeable" in what comes to the mediation estimates. The results of the test are provided in the Results section of the main manuscript.

*Testing differences in mediation effects between separate borderline personality disorder traits*

We additionally explored if separate psychopathological components of the borderline personality disorder were more important in mediating the effect of CM on suicidal behavior outcomes than the common construct underlying all the borderline criteria. According to previous work (6, 7) we formed four subcomponents constituting the DSM-IV diagnosis of BPD including an affective component (consisting of MSI items “increased anger”, “mood instability”, “feeling of emptiness”), an interpersonal component (MSI items “troubled relationships”, “avoidance of abandonment”), a behavioral component (MSI items “suicidal behavior”, “impulsivity”), and a cognitive component (MSI items “dissociative symptoms”, “distrustfulness”, “identity disturbance”). As the suicidality item in the behavioral component was omitted to avoid circularity, the single-item behavioral component was combined with the cognitive component to avoid potentially unreliable single-item measures. The sum scores of subcomponents were then scaled to same unit with the total score and mediation analyses were repeated.

*Comparing models based on TADS sub-scales versus overall TADS score*

In addition to using total TADS score, an omnibus test for differences among TADS subscales in predicting lifetime suicidal behavior (alternatively MSI) was conducted by comparing Akaike’s Information Criteria (AIC) (8) for models using sub-scales *versus* the main-scale as independent variables. We denote AIC of a sub-scale based model by AICsub and AIC of a main-scale model by AICmain, and let ΔAIC = AICsub ‒ AICmain. Lower AIC indicates more parsimonious model, ΔAIC < 0 meaning that the use of sub-scales improves the model over the original main-scale model.

*Missing data*

Only 10 patients had any missing values in the variables for mediation analyses (i.e., 3.5%). Therefore, we studied the complete (listwise deleted) data without missing-data imputations. Altogether 8 patients lacked more than two TADS items and were assigned as missing cases with respect to the total score, whereas for the others we computed the total score by averaging the non-missing items. Any missing items in TADS subscales were addressed as missing cases in pertinent analyses. Only one patient had more than one MSI items missing and was assigned as a missing case in the MSI total score, for the others, mean of the items formed the score. The BDI score was a mean of items, which was computed for all but the two patients with more than 2 missing items.

*Two direct and two indirect effects: counterfactual analysis*

The causal mediation package we applied has been developed from the counterfactual framework of causal inference (1). Assuming a random variable *Y* represents the outcome, the effect of changing the exposure (or “treatment”) variable *t* to a value *t + c* on the outcome can be computed as *Y*(*t + c*) – *Y*(*t*), where *Y*(*x*) refers to outcome of an individual when exposed at the level *x*. The typical problem of causal inference is that one observes only *Y*(*t + c*) or *Y*(*t*) for any given individual, not both, and therefore the within-individual effect is inaccessible (even in randomized experiments). Under certain conditions (e.g. randomization), the expected value E[*Y*(*t + c*) – *Y*(*t*)] can be computed, and we may regard *Y*(*t + c*) – *Y*(*t*) as a difference of *potential outcomes* if the exposure had been *t + c* *versus t*. Similarly, the mediator variable *M* could have been either *M*(*t + c*) or *M*(*t*). We chose *t + c* to be the maximum exposure and *t* the minimum exposure and computed the direct and indirect effects of the exposure on outcome based on this difference. However, the process still leaves us with two estimates for both the effects.

The direct effect is defined by holding the exposure constant for the mediator: *Y*(*t + c*, *M*(*x*)) – *Y*(*t*, *M*(*x*)), where *x* can be either *t + c* or *t* and the status of the mediator now also affects the outcome. The indirect effect is defined by holding the exposure constant with respect to its direct effect on the outcome instead of its effect on the mediator: *Y*(*x*, *M*(*t + c*)) – *Y*(*x*, *M*(*t*)). In principle, it could happen that the effect of mediator on the outcome depends on which “constant” counterfactual exposure one examines, *x = t + c* or *x = t*, which is why the “mediation” R package provides estimates for both the counterfactual values (i.e., stratifies for the constant factor). Of course, the stratified mediation effects can be averaged to obtain a total mediation effect, but we nevertheless plotted the non-averaged effects in the Figures for maximum transparency.

**References:**

1. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15(4):309-34.

2. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. Journal of Statistical Software. 2014;59(5):1-38.

3. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <https://www.R-project.org>.

4. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. New York: Cambridge University Press; 2007.

5. Good PI. Permutation, Parametric, and Bootstrap Tests of Hypotheses. 3 ed. New York, USA: Springer Science+Business Media, Inc; 2005.

6. Gunderson JG, Zanarini MC, Choi-Kain LW, Mitchell KS, Jang KL, Hudson JI. Family study of borderline personality disorder and its sectors of psychopathology. Arch Gen Psychiatry. 2011;68(7):753-62.

7. Reichborn-Kjennerud T, Ystrom E, Neale MC, Aggen SH, Mazzeo SE, Knudsen GP, et al. Structure of genetic and environmental risk factors for symptoms of DSM-IV borderline personality disorder. JAMA Psychiatry. 2013;70(11):1206-14.

8. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol Methods. 2012;17(2):228-43.