Writing about feelings, *and what is left of the effect*

Estimates of the health benefit of writing about upsetting events run from a medium effect in 1998 to a very small one later, which some analyses cannot tell from zero.

MITWA LAB, MUMBAI

IN PLAIN WORDS

In 1986 a small experiment had students write on four evenings, some about a trauma and some about trivial things. Only the group that wrote about both the facts and the feelings showed no rise in health-centre visits over the next six months, a difference that fell just short of the usual threshold for significance. Later studies pooled over a hundred experiments, and estimates of the average benefit run from medium to very small or indistinguishable from zero. In several studies, writing about distress made people feel worse for a while, and in one it left them more stressed than people who did not write. This review follows the numbers, says where the analyses disagree, and separates writing nobody answers from naming a feeling aloud and from talking with another person. It describes research and does not recommend the procedure.

Forty-six students and four evenings

Pennebaker and Beall (1986) asked 46 healthy undergraduates to write for 15 minutes on each of four consecutive evenings, after randomly assigning them to four groups. The controls (12 students) described trivial things, such as the room they were sitting in. The other three groups wrote about a personal trauma: its facts only (11 students), its feelings only (12), or both (11). Blood pressure and mood were measured before and after each session, and health-centre records gave each student's visits afterwards.

Straight after writing, the students who wrote about feelings, whether alone or together with the facts, had relatively higher blood pressure and worse mood than the facts-only and control groups. Over the following six months, visits to the health centre for illness rose in every group except the one that covered both facts and feelings, where they did not change. The test of that difference gave $p = .055$, just over the conventional .05 line, and the authors called their findings and their theory preliminary. They had students write rather than talk because they wanted to study disclosure without social feedback.

A medium effect from thirteen studies

Smyth (1998), read here as an abstract only, pooled the first wave of experiments. It reports significantly better health outcomes after the writing task in healthy participants, on self-rated physical health, well-being, physiological measures and overall functioning, but not on health behaviours. It also reports that writing raised distress right afterwards, an increase unrelated to later health. The abstract lists moderators, including whether the sample was college students, gender, duration, publication status and the wording of the instructions, but it gives no average effect size. Mogk et al. (2006) put it at $d = 0.47$ from 13 studies, 8 of them from the original research group, and call it significant and medium sized. Reinhold et al. (2018) quote 0.49 for the same analysis.

The statistic d is the difference between the writing and control groups in units of the standard deviation. Frattaroli (2006) notes that this analysis and a later one of clinical populations both assumed a single true effect common to all studies (a fixed-effects model), which limits how far the result can be generalised. A

2004 analysis of nine trials in clinical populations, as Mogk et al. describe it, found d = 0.19.

Two syntheses in 2006

Frattaroli (2006) gathered 146 randomised studies with 10,994 participants and used a random-effects model, which lets the true effect differ from study to study. The average correlation-type effect was r = .075, with a 95% interval of .051 to .098. She converts this to d = 0.15, and on her scale Smyth's analysis had averaged .257. She suggests that part of the gap comes from publication status, since 48% of her studies were unpublished against 23% in Smyth's, and unpublished studies tend to show smaller effects.

Frattaroli also argues for keeping the effect in view. In her account disclosing has virtually no costs, so a small average can still be worth having, and as an illustration (hers, not a measurement) an r of .075 would mean 54% of writers improving against 46% of controls in a hypothetical trial. Her average also includes studies run under poor conditions, some of them harmful (sessions shorter than 15 minutes, for example), and eight studies run mostly under favourable conditions, such as more sessions, privacy and specific instructions, averaged .200, which she describes as preliminary evidence.

Mogk et al. (2006) asked a narrower question. They kept only randomised trials in which the control group wrote about a neutral topic, each group had at least 10 people, and outcomes were measured at least four weeks after the last session. Of 216 studies identified, 42 met these criteria and 30 supplied usable data. Across 27 trials (one outlier removed) the average effect on health outcomes was g = 0.04 with a 95% interval from -0.08 to 0.15, where g is d with a correction for small samples. Psychological outcomes gave 0.01 and somatic outcomes 0.07. The one exception was an exploratory analysis of 16 studies of health behaviours such as sick days and doctor visits, which gave a significant 0.2 (interval 0.04 to 0.36). The authors suggest that the gap from Smyth may owe something to having 30 studies against 13 and to a conservative method. They also report that a review by Meads and colleagues (a 2003 report), which they learned of after their own study had been planned and conducted, reached the same conclusion.

By Frattaroli's own conversion her average of d = 0.15 sits at the upper edge of Mogk et al.'s interval, so on this arithmetic the two are not far apart in size. What differs is whether an effect that small can be told from zero, and that depends on which studies and outcomes are pooled.

A null trial and a subgroup that got worse

Niles et al. (2014) randomly assigned 116 healthy young adults, 59 to writing and 57 to a control topic, and analysed data from 113. The writers spent 20 minutes on each of four occasions on the most stressful or traumatic event of the previous five years. Writing had no significant effect on anxiety, depressive symptoms or physical symptoms. The authors tested several moderators (emotional expressiveness, emotional processing of the stressor and ambivalence over expressing emotion, plus gender and ethnicity), and only expressiveness, a trait measured by questionnaire, produced a significant moderating effect in the main analysis, and only on anxiety. Among writers, those high in expressiveness had a significant fall in anxiety at three months and those low in expressiveness a significant rise, while expressiveness did not predict change in anxiety among controls. The authors suggest that writing about a stressful event may be unsuitable for people who do not usually express emotions. That is a subgroup result from a single trial of 113 people, and the authors themselves describe earlier findings on whether low expressiveness predicts better or worse outcomes as inconsistent.

The authors also point out that their participants reported little anxiety, depression or illness to begin with, which leaves little room to improve. They add that among the studies in Frattaroli's analysis that used healthy college students, six found a significant benefit and 32 did not.

Reinhold et al. (2018) pooled 39 randomised trials (64 comparisons) in physically healthy adults without a diagnosis of post-traumatic stress disorder, looking only at depressive symptoms. Right after the last session there was a very small significant effect, g = -0.09, where a negative value means fewer symptoms. At the first follow-up the effect was -0.03 and not significantly different from zero, and a second follow-up was also not significant. Studies agreed with each other closely, and two checks for missing small null studies (a funnel plot and trim-and-fill) found none, so hidden null results do not obviously explain the finding. Larger effects went with more sessions and a more specific topic, although those are comparisons between studies and so cannot show cause. The authors concluded that brief, self-directed writing was not supported as a way of reducing depressive symptoms in this group of adults.

Change inside the writing group, a pandemic trial and a rival form of writing

Pavlacic et al. (2019), read as the authors' preprint, set the control group aside and pooled only the before-and-after change within the groups that wrote, using 202 effect sizes from 45 studies. For post-traumatic stress the random-effects estimate was d = 0.39 (95% interval 0.32 to 0.46), which the authors call small. A trim-and-fill correction, which imputed 19 missing studies, gave 0.26, other bias-adjusted methods gave anything from 0.12 to 0.60, and the authors report finding little evidence of publication bias. For post-traumatic growth it was 0.10 (0.02 to 0.17) and for quality of life -0.01 (-0.16 to 0.13). Without a comparison group, a fall in stress symptoms after writing cannot be told apart from the passage of time. That is this review's inference from the design; the authors do not make the point. They estimate the average statistical power of the growth studies at .15, and none reached the usual 80%.

Vukčević Marković et al. (2020) ran an online trial during the COVID-19 pandemic. They randomly allocated 120 adults to five writing sessions over two weeks or to treatment as usual, which the authors describe as a natural control. Primary outcome data came from 104 people, 48 writers and 56 controls. After adjusting for baseline scores, the writers did less well on overall psychological distress and on stress, and the groups did not differ on depression, anxiety, well-being or quality of life. Follow-up data came from 74 of the 120 people, and at one month there was no group difference on any outcome. The authors name a modest sample, a very mixed control group and a single writing instruction as limits, and they note that most writers still said taking part had been useful to them.

Guo (2023), also read as an abstract only, analysed 31 randomised trials with follow-up assessments (N = 4,012) in healthy and subclinical samples. Writing had a small but significant overall effect on symptoms of depression, anxiety and stress combined (g = -0.12, 95% interval -0.21 to -0.04), which emerged after a delay and was stronger when sessions were one to three days apart. Reinhold et al. found -0.03 at the first follow-up for depressive symptoms alone, so the two syntheses differ by roughly a tenth of a standard deviation, and they pool different outcomes. This review has not compared their study lists, so it cannot say why they differ.

Lai et al. (2023) compared expressive writing directly with positive writing in 24 randomised trials with 1,558 participants. In general-population samples, expressive writers reported more negative and less positive emotion than positive writers. On overall health, pain, physical symptoms and health visits the two kinds of writing did not differ, in analyses that were not split by population because each outcome had five or fewer studies. A comparison between two kinds of writing says nothing about whether either beats no writing at all.

Writing nobody answers, naming a feeling, talking to someone

The phrase "putting feelings into words" covers three situations, and the evidence for one does not carry over to the others.

Pennebaker and Beall (1986) built the writing paradigm itself to remove the listener. In Frattaroli's analysis, studies in which people disclosed alone in a room had larger effects than studies in which others were disclosing in the same room (r = .085 against .034). Spoken disclosure did no better than written, but the speaking was addressed to a tape recorder or to an experimenter who did not respond. Among six spoken studies, speaking to an experimenter averaged r = .106 and speaking to a recorder .048, a difference that was not significant.

Naming a feeling in the laboratory is a different case. Kircanski et al. (2012) randomly assigned 88 adults fearful of spiders to four groups of 22. During repeated exposure to a live spider, one group spoke a sentence naming the spider and their own emotion, and the other groups reframed the situation, distracted themselves or received no instruction. At a test one week later the naming group showed a larger fall in skin conductance, measured from the test straight after the sessions, than each other group (d = 0.85, 0.74 and 0.64), but it did not differ in self-reported fear. Right after the sessions, skin conductance did not differ significantly between groups. A spoken sentence during exposure to a feared object, measured by sweat response, is a different thing from a written account over several days, measured by outcomes such as health-centre records or symptom questionnaires, so a result from one says little about the other.

Talking to another person is the third case. The Cochrane review of single-session debriefing (Rose et al., 2002 version) covered 15 trials of a one-off session in which people exposed to a trauma within the previous month were encouraged to recall the event and were reassured that their reactions were normal. Most trials were of poor quality. Debriefing did not prevent post-traumatic stress disorder or reduce distress compared with control, and one trial

found a significantly higher risk at one year (odds ratio 2.51, 95% interval 1.24 to 5.09). That is one structured, one-off kind of talk, and it says little about the many conversations of ordinary life, including a conversation with a friend who answers, which none of the trials above test.

A reply, whether from a person or from software, adds a responder, which the original writing design set out to leave out. Frattaroli excluded 14 studies in which experimenters gave oral or written feedback, because that closely resembles psychotherapy, and Reinhold et al. cite another group's analysis of nine therapist-delivered writing studies in samples with post-traumatic stress disorder (g = 0.81). The studies described here were not designed to test writing that is answered by a companion or a chatbot.

What is not yet known

- The reason Guo (2023) finds a small delayed effect where Reinhold et al. (2018) find none at follow-up. The abstracts show differences in outcomes, but this review did not examine whether the two share trials or how the result changes if they are re-run on the same ones.
- Whether the favourable conditions Frattaroli (2006) singled out, such as more sessions, privacy and specific instructions, produce an effect larger than small. Her figure of .200 comes from eight studies, which she calls preliminary evidence.
- Who is made worse. The rise in anxiety among low-expressive writers comes from one trial of 113 people, in a literature its authors call inconsistent, and the worse stress among writers during the pandemic comes from one trial of 104.
- Whether the before-and-after changes in post-traumatic stress that Pavlacic et al. (2019) pooled would survive a comparison with control groups.
- Whether anything above applies when the writing receives a reply. The original paradigm was designed without one. The sources read here mention writing with a responder only in passing (Frattaroli's exclusion of studies with experimenter feedback, and the g = 0.81 result for therapist-delivered writing cited by Reinhold et al.) and describe no trial of writing answered by a companion or chatbot.
- This review is not a systematic search. Smyth (1998) and Guo (2023) were read as abstracts only, so their moderator analyses and study lists are not examined, and the 2004 clinical-populations analysis and the review by Meads and colleagues are known here only through the accounts of others.

References

Frattaroli, J. (2006). Experimental disclosure and its moderators: a meta-analysis. Psychological Bulletin, 132(6), 823-865. doi.org/10.1037/0033-2909.132.6.823 (read as a copy at gwern.net/doc/psychology/2006-frattaroli.pdf).

Guo, L. (2023). The delayed, durable effect of expressive writing on depression, anxiety and stress: a meta-analytic review of studies with long-term follow-ups. British Journal of Clinical Psychology, 62(1), 272-297. pubmed.ncbi.nlm.nih.gov/36536513/ (abstract only).

Kircanski, K., Lieberman, M. D., and Craske, M. G. (2012). Feelings into words: contributions of language to exposure therapy. Psychological Science, 23(10), 1086-1091. pmc.ncbi.nlm.nih.gov/articles/PMC4721564/

Lai, J., Song, H., Wang, Y., Ren, Y., Li, S., Xiao, F., Liao, S., Xie, T., and Zhuang, W. (2023). Efficacy of expressive writing versus positive writing in different populations: systematic review and meta-analysis. Nursing Open, 10(9), 5961-5974. pmc.ncbi.nlm.nih.gov/articles/PMC10415981/

Mogk, C., Otte, S., Reinhold-Hurley, B., and Kröner-Herwig, B. (2006). Health effects of expressive writing on stressful or traumatic experiences - a meta-analysis. Psycho-Social Medicine, 3, Doc06. pmc.ncbi.nlm.nih.gov/articles/PMC2736499/

Niles, A. N., Haltom, K. E., Mulvenna, C. M., Lieberman, M. D., and Stanton, A. L. (2014). Effects of expressive writing on psychological and physical health: the moderating role of emotional expressivity. Anxiety, Stress, and Coping, 27(1). pmc.ncbi.nlm.nih.gov/articles/PMC3830620/

Pavlacic, J. M., Buchanan, E. M., Maxwell, N. P., Hopke, T. G., and Schulenberg, S. E. (2019). A meta-analysis of expressive writing on posttraumatic stress, posttraumatic growth, and quality of life. Review of General Psychology, 23(2), 230-250. Read as the authors' preprint, which carries a slightly different title: osf.io/u98cw

Pennebaker, J. W., and Beall, S. K. (1986). Confronting a traumatic event: toward an understanding of inhibition and disease. Journal of Abnormal Psychology, 95(3), 274-281. doi.org/10.1037/0021-843X.95.3.274 (read as a copy of the published article).

Reinhold, M., Bürkner, P.-C., and Holling, H. (2018). Effects of expressive writing on depressive symptoms: a meta-analysis. Clinical Psychology: Science and Practice, e12224. doi.org/10.1111/cpsp.12224 (read as the authors' copy at paul-buerkner.github.io/publications/pdf/2018__Reinhold_et_al__Clinical_Psychology_Science_and_Practice.pdf).

Rose, S. C., Bisson, J., Churchill, R., and Wessely, S. (2002). Psychological debriefing for preventing post traumatic stress disorder (PTSD). Cochrane Database of Systematic Reviews, 2002(2), CD000560. pmc.ncbi.nlm.nih.gov/articles/PMC7032695/

Smyth, J. M. (1998). Written emotional expression: effect sizes, outcome types, and moderating variables. Journal of Consulting and Clinical Psychology, 66(1), 174-184. pubmed.ncbi.nlm.nih.gov/9489272/ (abstract only).

Vukčević Marković, M., Bjekić, J., and Priebe, S. (2020). Effectiveness of expressive writing in the reduction of psychological distress during the COVID-19 pandemic: a randomized controlled trial. Frontiers in Psychology, 11, 587282. pmc.ncbi.nlm.nih.gov/articles/PMC7683413/