Study Protocol and Statistical analysis plan

Can Massage During One Year Improve Health in Health-care Providers Working in Hospital

NCT: 05555082

Date: 2025-10-17

Study design.

This study used an embedded mixed methods pre-post design, where qualitative data were collected at the final follow-up to complement and help interpret quantitative finding [1]. A mixed-methods approach combines quantitative and qualitative data to explore a phenomenon more comprehensively, allowing statistical trends (quantitative) to be explained through qualitative insights and qualitative findings to complement or expand quantitative results [1]. The embedded design is flexible and adaptable, making it valuable for complex research questions where both numerical and narrative insights are needed. Quantitative data were prioritized to appraise limited efficacy, defined under Bowen's feasibility framework as a preliminary within group assessment of the direction and magnitude of change, and to assess acceptability indicators including recruitment , adherence, and participant satisfaction [2] . Qualitative data were collected at 12 months to deepen understanding of participants' experiences [1] . Integration occurred during analysis by comparing qualitative themes with quantitative outcomes to examine convergence and divergence, with details provided in the Results and the Discussion.

Participants

While the literature reviewed in the introduction refers to (HCW) broadly, our sample was limited to registered and assistant nurses. This focus reflects the reality of the inpatient ward setting, where these professionals are most exposed to physically and emotionally demanding conditions, including shift work and close patient contact. Thus, registered and assistant nurses working in a medical ward at a Swedish hospital were eligible to participate in the study. To be included, participants needed to work in patient care at the ward for at least 50% of full-time hours**.** Approximately 130 nurses were estimated to meet the eligibility criteria, and we planned to recruit 50 nurses. The sample size was pragmatically based on operational capacity and resource availability, for example available funding only allowing us to offer a limited number of massage sessions. In addition, the number was considered reasonable to assess feasibility parameters such as recruitment yield, adherence, and acceptability, in line with Bowen et al. (2009) [2]. While not statistically powered, the target was selected to support interpretation of feasibility outcomes. Recruitment was scheduled for a one-month period (November–December 2022) before the planned start of the intervention in January 2023. To raise awareness ahead of this short recruitment window, information about the upcoming study was shared with staff during routine staff meetings prior to the formal registration period. Recruitment was then carried out through an advertisement on the clinic's

website and an email sent to all healthcare staff at the clinic. The recruitment email was directed specifically to the clinic's nursing staff, in accordance with internal mailing list structures. Other professional groups, such as physicians, physiotherapists, or social workers, were not contacted and were not eligible to participate. Both the advertisement and the email provided detailed information about the study, including the maximum number of participants (50), and included a link to a registration form for those interested. The form collected participants' names, phone numbers, and email addresses.

Individuals who registered and meet the inclusion criteria were emailed and invited to meet with the research nurse. During the appointment, participants had the opportunity to ask the nurse additional questions about the study and were asked to provide written consent for participation. At the visit baseline measures were obtained, including blood pressure, height, and weight, together with the baseline questionnaires.

The intervention

The participants received classic massage in 30-minute sessions approximately once a month over the course of one year (i.e., 2023). This duration was based on the standard hands-on treatment time offered by the certified massage therapists contracted for the study. Massage sessions were scheduled outside of working hours in agreement between each participant and the therapist and took place in a quiet and private room at a massage clinic to support relaxation.

Two certified massage therapists were engaged for the study, and participants were evenly divided between them. Each participant received all their sessions from the same therapist to promote continuity and individual adaptation of the treatment. Both therapists were certified in classic Swedish massage and had 15 and 21 years of experience operating private massage practices. The treatment followed the principles of classic massage, including kneading, gliding, and tapping techniques, and primarily targeted areas commonly affected by work-related strain, such as the back, shoulders, and neck. The depth and focus of the massage were tailored to individual needs through ongoing dialogue with each participant. Because ward operations and summer vacations were expected to potentially reduce availability in June and July, the protocol allowed rescheduling across adjacent months; no blanket pause was planned. The target total exposure per participant was 10 to 12 sessions within twelve months, depending on individual scheduling. No exclusion criteria were set based on exact month-to-month spacing.

Data collection

Data was collected at three time points: baseline, 6 months, and at the end of the massage intervention after 12 months. Baseline data were collected at the initial appointment with the research nurse. Prior to this appointment, participants received a package of questionnaires to complete at home, which they returned during the visit. This procedure allowed participants time to reflect on the questions while ensuring that the completed forms could be checked for completeness or clarified during the appointment. These questionnaires covered background variables and outcome variables of interest (i.e., stress, mental health, sleeping problems, physical activity, HR-QoL and work ability). Participants were provided with a stamped envelope to return the completed questionnaires, either by handing them to the research nurse or by mailing them to the project leaders at the clinic, PJ or LE. At the 6-month follow-up, participants completed the same set of questionnaires as at baseline which included questions about the number of massage sessions they attended, as well as questions about stress, mental health, sleep problems, HR-QoL, and work ability. At the end of the intervention at 12 months (December 2023–January 2024), participants again completed the same set of questionnaires, along with additional items on physical activity, satisfaction, and open-ended questions. On this occasion, they also rated their physical activity and scored their satisfaction with the massage on a scale, as well as responded to open-ended questions, allowing for the collection of both quantitative and qualitative data at 12 months-follow-up.

Feasibility outcomes

Limited efficacy

The quantitative outcome measures were collected to explore such early signals of effect, assess the relevance and sensitivity of the instruments used, and inform the design of a future fully powered trial. This does not imply low effectiveness, but rather a structured exploration of whether further testing is warranted [2].

Quantitative data collection

Stress

Stress levels were evaluated using the Perceived Stress Scale-10 (PSS-10)[3]. The questionnaire consists of 10 items that assess how unpredictable, uncontrollable, and overloaded respondents find their lives. Participants rate each item on a 5-point Likert scale, ranging from 0 (never) to 4 (very often), reflecting their feelings and thoughts during the past month. The scores for the PSS-10 range from 0 to 40, with higher scores indicating higher

perceived stress. The PSS-10 has been validated in Swedish and is considered a reliable indicator of perceived stress [4].

Mental Health

We employed three different questionnaires to assess various aspects of mental health: the Patient Health Questionnaire-9 (PHQ-9)[5] for depressive symptoms, the General Anxiety Disorder Scale-9 (GAD-7)[6] for anxiety, and the Karolinska Exhaustion Disorder Scale (KEDS) [7] for exhaustion.

PHQ-9 consists of nine questions aligned with the diagnostic criteria for major depressive disorder in the DSM-IV[5]. Each question is scored on a scale from 0 (not at all) to 3 (nearly every day), considering the frequency of symptoms over the past two weeks. The total score, ranging from 0 to 27, helps gauge the severity of depression, with higher scores indicating more severe depressive symptoms. The cut-off scores for PHQ-9 are as follows: Minimal depression ranges from 0 to 4, mild depression from 5 to 9, moderate depression from 10 to 14, moderately severe depression from 15 to 19, and severe depression from 20 to 27. A ≥20% reduction from baseline is typically used to indicate a clinically meaningful improvement in depressive symptoms [5].

GAD-7 [6] comprises seven questions, each rated on a scale from 0 (not at all) to 3 (nearly every day), reflecting the frequency of anxiety symptoms over the past two weeks. The total score ranges from 0 to 21, with higher scores indicating greater anxiety severity. The cut-off scores for GAD-7 are as follows: minimal anxiety ranges from 0 to 4, mild anxiety from 5 to 9, moderate anxiety from 10 to 14, and severe anxiety from 15 to 21. A 4-point reduction has been suggested as the threshold for minimum clinically important difference (MCID)[8].

Sleep problems

The Insomnia Severity Index (ISI) [9] assessed the severity of both nighttime and daytime components of insomnia. It consists of seven items that evaluate the nature, severity, and impact of insomnia over the past two weeks. Each item is rated on a 5-point Likert scale ranging from 0 (no problem) to 4 (very severe problem), with the total score ranging from 0 to 28. Higher scores on the ISI indicate greater severity of insomnia symptoms. The cut-off scores for interpreting the ISI are as follows: scores from 0 to 7 indicate no clinically significant insomnia; scores from 8 to 14 suggest subthreshold insomnia; scores from 15 to 21 reflect clinical insomnia of moderate severity; and scores from 22 to 28 indicate severe clinical insomnia. For ISI, two MCID thresholds have been proposed based on different outcome anchors. A 6-point reduction is commonly used to indicate clinically meaningful

improvement in individuals with primary insomnia [10]. However, smaller changes around 2.5 points have also been associated with meaningful improvements in functioning and mood in non-clinical populations [10]. In this study, we report the proportion of participants meeting both thresholds to provide a range of interpretive perspectives, recognizing the non-clinical nature of our sample.

Health-related quality of life

The RAND-36 was used to measure HR-QoL and includes 36 items that measure eight health domains: physical functioning, role limitations due to physical health problems, role limitations due to emotional problems, energy/fatigue, emotional well-being, social functioning, pain, and general health perceptions. An additional item assesses the change in health the last 12 month and is labelled health transition. Each domain is scored on a scale from 0 to 100, with higher scores indicating better health outcomes [11].

Data analysis

In the quantitative analysis, descriptive statistics were used to summarize participant characteristics, background variables, recruitment, adherence, and satisfaction rates. Categorical data were reported as counts and percentages, while continuous variables were presented as means with standard deviations (SD) or medians with interquartile ranges, depending on the data distribution. To evaluate the limited efficacy of the intervention, Linear Mixed Models (LMM) were applied to data collected at baseline, 6 months, and 12 months. LMM was chosen for its ability to handle repeated measures and missing data, with time treated as both a fixed and repeated effect, an unstructured covariance type, and maximum likelihood estimation [12]. A paired sample t-test was conducted to analyze changes in physical activity between baseline and 12 months. In line with recommendations for feasibility studies by Bowen et al. (2009) [2], effect sizes were prioritized to assess limited efficacy and P-values were reported for transparency but were not used to infer significance, as feasibility studies are not powered for hypothesis testing. For each outcome, we reported means, SDs, change scores with 95% confidence intervals (CI), p-values and effect sizes (Cohen's *d*), interpreted using conventional thresholds where 0.20–0.49 indicates a small effect, 0.50–0.79 a moderate effect, and ≥0.80 a large effect.

Limited efficacy was treated as an exploratory within group appraisal under Bowen's framework. We summarised the pattern across domains rather than any single measure and flagged a limited efficacy signal when at least three target domains showed a Cohen's d of

0.20 or greater at 6 or 12 months. We considered durability to be supported when these changes were maintained or further improved at 12 months. These parameters describe feasibility stage signals and are not hypothesis tests or pass or fail rules. All statistical analyses were performed using IBM SPSS version 25.0.

References

1. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research: SAGE Publications 2011.
2. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. *Am J Prev Med* 2009;36(5):452-7. doi: 10.1016/j.amepre.2009.02.002 [published Online First: 2009/04/14]
3. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24(4):385-96. [published Online First: 1983/12/01]
4. Nordin S, Palmquist E, Nordin M. Psychometric evaluation and normative data for a Swedish version of the Patient Health Questionnaire 15-Item Somatic Symptom Severity Scale. *Scand J Psychol* 2013 doi: 10.1111/sjop.12029 [published Online First: 2013/01/09]
5. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606-13. [published Online First: 2001/09/15]
6. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;166(10):1092-7. doi: 10.1001/archinte.166.10.1092 [published Online First: 2006/05/24]
7. Besèr A, Sorjonen K, Wahlberg K, et al. Construction and evaluation of a self rating scale for stress-induced exhaustion disorder, the Karolinska Exhaustion Disorder Scale. *Scand J Psychol* 2014;55(1):72-82. doi: 10.1111/sjop.12088 [published Online First: 2013/11/19]
8. Toussaint A, Hüsing P, Gumz A, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). *J Affect Disord* 2020;265:395-401. doi: 10.1016/j.jad.2020.01.032 [published Online First: 2020/02/25]
9. Morin CM, Belleville G, Belanger L. Validation of the insomnia severity index. *Sleep* 2006;29:A258-A59.
10. Yang M, Morin CM, Schaefer K, et al. Interpreting score differences in the Insomnia Severity Index: using health-related outcomes to define the minimally important difference. *Curr Med Res Opin* 2009;25(10):2487-94. doi: 10.1185/03007990903167415 [published Online First: 2009/08/20]
11. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ* 1993;2(3):217-27. doi: 10.1002/hec.4730020305 [published Online First: 1993/10/01]
12. Molenberghs G, Verbeke G. A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling* 2001;1(4):235-69. doi: 10.1177/1471082X0100100402