Integrative Cancer Therapies
Volume 19: 1–7
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1534735420903801
journals.sagepub.com/home/ict
SAGE

Interrater Reliability of Dynamic Muscle Testing After Breast Cancer Surgery in Women at High Risk of Lymphedema: To Improve Quality in Clinical Practice

Tina Korsholm-Rosfort, BSc[1], Simone Larsen, BSc[1], Thomas Aagaard, MSc[2,3,4], Gunn Ammitzbøll, MSc, PhD[5,6], and Susanne O. Dalton, MD, PhD[5,6]

Abstract

Background: The purpose of this study is to determine the interrater reliability of dynamic muscle tests in the early rehabilitation phase in women after breast cancer surgery with axillary lymph node dissection (ALND) based on the "preventive intervention against lymphedema after breast cancer" (LYCA) randomised controlled trial. **Methods:** Fifteen women treated with breast cancer surgery including ALND were recruited from participants in the LYCA trial. In this interrater reliability study, women were tested in 4 dynamic muscle tests by 2 physiotherapists at a Capital Hospital in Denmark. Intraclass correlation coefficients (ICCs) with 95% confidence intervals (CIs) was used to assess the relative reliability between raters. A Bland-Altman plot and limits of agreement were calculated to describe the absolute reliability. **Results:** All 6 subtests displayed high interrater reliability. ICC values were: leg press 0.96 (95% CI = 0.87-0.99), elbow flexion (contralateral) 0.94 (0.83-0.98), elbow flexion (affected arm) 0.93 (0.80-0.98), elbow extension 0.80 (0.41-0.93), shoulder abduction (contralateral) 0.89 (0.68-0.96), and shoulder abduction (affected arm) 0.91 (0.74-0.97). Cumulated interrater reliability for the test battery was very high (ICC = 0.99, 95% CI = 0.990-1.0). The absolute reliability of this study was considered high, and the absence of large shifts between mean and the line of 0 difference suggest no systematic bias that could influence clinical interpretation. **Conclusions:** The dynamic muscle tests evaluated in this study had high interrater reliability and can be used reliably in women in the postoperative phase after breast cancer surgery with ALND.

Keywords

reproducibility of results, neoplasms, lymphedema, breast cancer lymphedema, breast neoplasms, resistance training, interrater reliability

Submitted August 9, 2019; revised December 16, 2019; accepted December 31, 2019

Introduction

Breast cancer is the most frequently diagnosed cancer type in women worldwide, affecting more than 1.5 million women each year[1] and representing 25% of all female cancer cases.[2] The 5-year survival rate of breast cancer in many countries with advanced medical care is estimated to 80% to 90%; in Denmark, it is 86%; and by the end of 2014, 62 150 Danish women were living with the diagnosis.[3]

Breast cancer treatment often involves a surgical intervention and various combinations of adjuvant chemotherapy, radiotherapy, and/or hormone treatment. All aspects of the disease and treatment protocol stress the lymphatic system, but in particular, axillary lymph node dissection

[1]Department of Physiotherapy and Occupational Therapy, Faculty of Health and Technology, University College Copenhagen, Denmark
[2]Department of Physio- and Occupational Therapy, Holbaek Hospital, Denmark
[3]Department of Orthopedic Surgery, Holbaek Hospital, Denmark
[4]Department of Orthopedic Surgery, Zealand University Hospital, Denmark
[5]Danish Research Center for Equality in Cancer (COMPAS), Zealand University Hospital, Naestved, Denmark
[6]Survivorship and Equality in Cancer, Danish Cancer Society Research Center, Copenhagen, Denmark

Corresponding Author:

Simone Larsen, Smedelundsgade 36, 1 th, Holbaek, 4300, Denmark.
Email: simonelarsen.reliability.study@outlook.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

(ALND) and radiotherapy increase the risk of breast cancer–related lymphedema (BCRL).[4] Focus in cancer rehabilitation postsurgery has been on increasing range of movement and function of the upper limb. Historically, guidelines and recommendations after breast cancer surgery have been followed to avoid loading beyond 5 to 15 kg,[5,6] based on the belief that excessive loading could trigger BCRL. In previous research on progressive strength training and risk of lymphedema after breast cancer, standard practice has been to start exercise at a minimal or moderate load and increase gradually by the smallest increment available.[7-10] To our knowledge, only one comparable study on postsurgery breast cancer patients has utilized a strength test to determine the individual starting point for progressive strength training, but with no evidence of reliability of the strength tests used.[11] Arguments for not performing strength tests in individuals after breast cancer surgery could involve caution due to healing concerns, though this is not supported in the literature. Furthermore, knowledge about lymphedema risk has been insufficient to reassure clinicians that lymphedema would not be triggered by a bout of maximal muscle exertion. However, 2 recent studies have investigated the effect of a single session of different load strength exercises concluding that extracellular fluid in the arm is not increased, neither from high-load nor from moderate-load exercise.[12,13] It can therefore be assumed that maximal muscle testing in this high-risk population is safe with regard to BCRL development and status.

To make maximal use of the benefits of progressive strength training, it is necessary to tailor the strength training to the individual patients' levels of strength by initial and repeated testing throughout the course of rehabilitation. With recent advances in research documenting the safety of maximal strength testing, it is essential to examine the reliability of these strength tests so they can be used for scientific and rehabilitation purposes. When exploring the interrater reliability, 2 important and non-interchangeable aspects can be considered; the level of resemblance between test measurements and whether they keep their order in the hierarchy (relative reliability), and the level of variation between the value of the test results by each tester (absolute reliability).

The objective of this study was to examine the interrater reliability of dynamic muscle tests in the early rehabilitation phase in women at high risk of BCRL who have undergone breast cancer surgery with ALND. To our knowledge, this is the first published study to investigate the interrater reliability of maximal dynamic muscle tests in this specific group of patients.

Methods

Participants and Setting

In this study, we recruited participants from an ongoing randomized controlled trial, the preventive intervention against LYmphedema after breast CAncer (LYCA) study, which examined the effect of progressive strength training on the development of BCRL in the first year after surgery. Participants were women who had undergone breast cancer surgery and ALND. The intervention group received 20 weeks of supervised exercise in teams followed by 30 weeks of self-managed exercise. The control group received usual care. Further details of the LYCA study design have been described elsewhere.[14,15]

Participants included in the present study were recruited by convenience sampling among women participating in LYCA from the 2 hospitals in the Capital Region, and from both the control and intervention groups. In order to test the interrater reliability, adjustments to the LYCA protocol were made to ensure blinding based on the recommendations for diagnostic procedures in manual/musculoskeletal medicine by the International Academy of Manual/Musculoskeletal Medicine.[15,16] Adjustments were (1) participants were blinded to the weights being lifted during testing and (2) raters made the choice of starting weight. Furthermore, all participants were instructed not to inform the raters about results previously achieved in the LYCA study.

Sample Size

According to de Vet et al,[17] an interclass correlation coefficient (ICC) level of 0.8 with a 95% confidence interval (CI) $\pm$ 0.2 requires a sample size of at least 13 participants with 2 repeated measurements with regard to reliability.[16] Based on this, and with the resources at hand in this project, we aimed toward a total sample size of 18 with a sample size down to $n = 13$ as statistically acceptable.

Raters

Raters involved in this study had previous experience with the test protocol used (Appendix A, available online); 1 physiotherapist with 18 months of experience in the LYCA study performing the test protocol, and 1 physiotherapist acting as coordinator and developer of the study protocol in the LYCA study although with no practical experience with testing the protocol in patients. Both raters received the protocol before the first test day and were thoroughly instructed in blinding procedures.

Equipment

Equipment used in this study included dumbbells, a triceps resistance machine with a seated starting position, and a leg press resistance machine. Furthermore, tape for markings on the wall and a stopwatch were used for the execution of the dynamic muscle tests.

Procedure

All participants went through a series of four 7-RM (repetition maximum) dynamic muscle tests within a total period of

20 minutes. Testing was carried out in a fixed order of paired tests to minimize bias from muscle fatigue: (1) elbow flexion and legpress, and (2) shoulder abduction and elbow extension. Participants rested for 10 minutes between the 2 series of tests by each rater. The order in which raters executed testing varied so that rater 1 and rater 2 both would act as first rater. The participants were blinded to all outcomes during testing, and the results were later reported by the clinicians.

All tests were performed according to a detailed test protocol (Online Resource 1, available online), with well-described criteria for a repetition to be accepted. In each test, 7 RM was defined as the maximum weight in whole numbers of kilograms the participants could lift 7 times in good quality (evaluated by the rater) throughout full predefined range of motion. The maximum test results were documented in a test form and kept hidden from participants and raters until the end of testing.

Statistical Analysis

The data used in this study were ratio-interval data and has therefore been analyzed using ICC, limits of agreement (LOA), and Bland-Altmann.

Interrater reliability was calculated using the ICC model 2.2 for each dynamic muscle test. ICC is a correlation coefficient that indicates the level of resemblance between measurements. An ICC level close to 1 indicates a high level of similarity, whereas an ICC level close to 0 means no similarity. ICCs were evaluated according to the following standards reported by Portney and Watkins[17,18]: poor reliability <0.50, moderate reliability $\approx$ 0.50 to 0.75, and high reliability >0.75. ICC scores were calculated for each individual subtest, as well as cumulated ICC for the test battery.

In the assessment of absolute reliability, LOA and Bland-Altman plots were used to determine interrater agreement and detect any potential systematic bias. In the Bland-Altman plots, variations between the absolute values of the measurements are detected. It is recommended that 95% of the data points in the plots should lie within ± 2 standard deviations (SDs) of the mean difference.[18,19] A large shift of mean away from the line of 0 difference would imply occurrence of bias in the results. The 95% LOA depends on the size of the sample, a small sample size giving a larger variation between upper and lower limit.[19,20] Data were analyzed using IBM SPSS Statistics Version 25. Estimates were regarded as statistically significant with $P < .05$.

Ethical Considerations

All participants signed an informed consent form before any testing was carried out. The study was approved by the National Committee on Health Research Ethics under the approval for the LYCA study (Protocol Number: H-16019587).

Results

Participants

All women included had undergone breast cancer surgery with ALND and participated in the LYCA study in their first year after surgery. A total of 55 women out of the 158 women included in the LYCA study were contacted by phone, and 15 women accepted participation in the present study (Figure 1). The mean age of participants was 56 years (SD = 8.8), and mean time since surgery was 61 weeks (SD = 19). Three participants (20%) had been allocated to the control condition in the randomized controlled trial.

Relative Reliability

We found high relative reliability[20,21] with ICC values for all subtests ranging from 0.80 to 0.96 (Table 1). The overall cumulated ICC values for the test battery was 0.99 (CI = 0.990-1.00) indicating very high interrater reliability.

Absolute Reliability

Data presented in Table 2 show results for absolute reliability for each of the 6 tests. We found the largest intervals between upper and lower LOA for leg press and elbow extension, with a mean difference between measurements for the 2 raters of 3 ± 21 kg and -6 ± 12 kg, respectively (Table 2). This corresponds to a mean deviation of 1 increment in the leg press machine (10 kg) and 1 increment in the elbow extension machine (10 kg). With a 95% certainty, the estimate will lie between −20 kg and 25 kg for leg press and between −28 kg and 17 kg for elbow extension between the 2 raters. In the Bland-Altman plots (Online Resource 2, available online), we looked for variations between the measurements. Since the variations between measurements from our 2 raters never exceeded 1 increment, the absolute reliability of this study is considered to be high. We found broad intervals between the limits of agreement; however, due to the absence of large shifts between mean and the line of 0 difference, we found no evidence of systematic bias that could influence clinical interpretation. When comparing mean values of rater 1 with mean values of rater 2 for each of the 6 tests in Table 2, there seems to be no systematic bias between the raters.

Discussion

Major Findings and Strengths

This study investigated the interrater reliability of the 7-RM tests used in a population of women who had undergone breast cancer surgery with ALND. We found that the

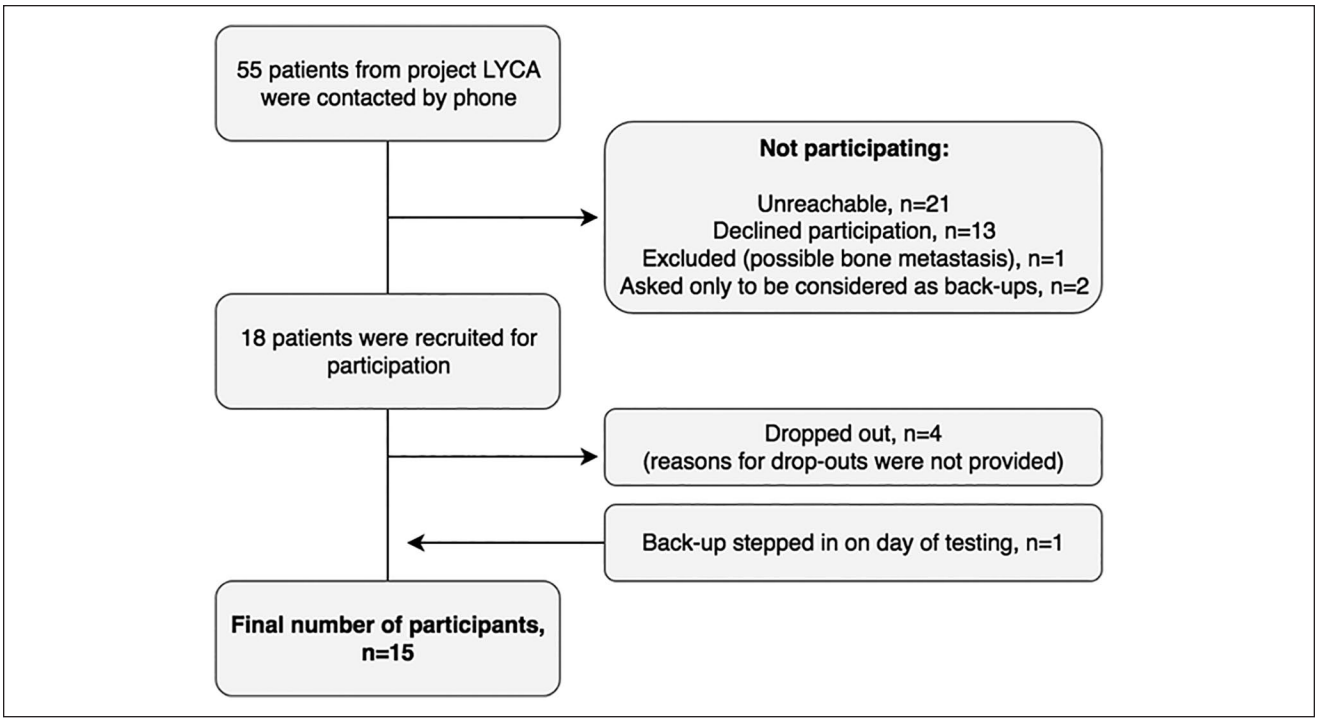

Figure 1. Flowchart showing the inclusion process in the study of interrater reliability of dynamic muscle testing in women after breast cancer surgery with axillary lymph node dissection (ALND), the preventive intervention against LYmphedema after breast CAncer (LYCA) study, East Denmark, 2015 to 2018.

Table 1. Relative Reliability for 7–Repetition Maximum Tests Described by ICC and 95% CIs in 15 Women After Breast Cancer Surgery With ALND, the LYCA Study, East Denmark, 2015 to 2018.

Exercise	ICC	95% CI
Leg press	0.96	(0.87-0.99)
Elbow flexion	0.94	(0.83-0.98)
Elbow flexion AA	0.93	(0.80-0.98)
Elbow extension	0.80	(0.41-0.93)
Shoulder abduction	0.89	(0.68-0.96)
Shoulder abduction AA	0.91	(0.74-0.97)
Total for the test battery	0.99	(0.99-1.00)

Abbreviations: ICC, intraclass correlation coefficient; CI, confidence interval; ALND, axillary lymph node dissection; LYCA, preventive intervention against LYmphedema after breast cancer; AA, affected arm.

relative reliability described by the ICC values was high for the 6 subtests. Furthermore, the absolute reliability assessing the agreement between raters was also found to be high, and our results indicate no systematic bias was present in the 6 subtests.

Strengths of this study include the translatability of tests to clinical practice, and the avoidance of handheld dynamometry, which has lower reliability and is often not available in the clinical setting. A further strength is the double-blinded design where patients were blinded to the maximum load they reached in the muscle testing, and the raters were blinded to results on each patient. On the other hand, the design could be argued to introduce a risk of fatigue, as the raters were granting participants an unlimited number of trials to achieve the 7-RM determination. This risk was minimized by maintaining an intermission of 10 minutes between participants' first and second rounds of tests. Conducting the test sessions on 2 separate days might have reduced the risk of fatigue even further, but same-day testing was chosen with the argument that the day-to-day variation in current shape in this population is large due to adjuvant therapies, which would compromise repeatability to a greater extent.[21,22]

Meaning and Importance of the Findings

The establishment of high interrater reliability in this exact patient group is crucial for future implementation of progressive strength training in a postsurgery rehabilitation program. The validation of these tests will act as methodological pillars in a process where pervious notions and traditions are questioned, and the safety of high-load exercise for patients at risk for BCRL is being established. It has been suggested that a slow increase in the maximum work capacity will give these women a higher threshold for when

Table 2. Mean Values for the Difference Between the 2 Raters, and Lower and Upper LOA in 15 Women After Breast Cancer Surgery and ALND, the LYCA Study, East Denmark, 2015 to 2018.

Exercise	Mean Rater 1 ± SD[a]	Mean Rater 2 ± SD[b]	Mean Difference (±SD)[a]	Lower LOA[b]	Upper LOA[c]
Leg press	112.00 ± 24.842	113.33 ± 31.547	2.67 ± 11.63	−20.13	25.46
Elbow flexion	6.60 ± 1.404	6.73 ± 1.163	−0.27 ± 0.59	−1.43	0.90
Elbow flexion AA	6.73 ± 1.438	6.40 ± 1.352	0.07 ± 0.70	−1.31	1.45
Elbow extension	58.67 ± 15.055	63.67 ± 13.558	−5.67 ± 11.63	−28.46	17.13
Shoulder abduction	4.53 ± 0.834	4.40 ± 0.910	−0.53 ± 0.52	−1.55	0.48
Shoulder abduction AA	4.40 ± 0.910	4.33 ± 0.976	−0.47 ± 0.52	−1.48	0.55

Abbreviations: LOA, limits of agreement; ALND, axillary lymph node dissection; LYCA, preventive intervention against LYmphedema after breast cancer; SD, standard deviation; AA, affected arm.
[a]Mean (kg) for the difference (rater 1 − rater2) ± SD (kg).
[b]Lower limit (kg) for 95% confidence interval (LOA): Mean − (SD·1.96).
[c]Upper limit (kg) for 95% confidence interval (LOA): (SD·1.96) + Mean.

normal everyday activities become an overload; and therefore, the risk of developing or worsening BCRL due to inflammatory response will be decreased.[6] Valid strength measurement methods have their importance in this setting, where the individual level and adaptation to exercise needs to be addressed for strength training to be effective.[22,23]

The safety of high-load exercise is further reinforced by a recent study by Bloomquist et al, where the acute lymphatic response following heavy-load resistance exercise did not exceed the response following low-load exercise and stresses that heavy-load exercises do not increase extracellular fluid in women with axillary node dissection at high risk of developing BCRL.[23,24] The 21 participants training with loads ranging from 85% to 90% of 1 RM for 5 to 8 repetitions showed no exacerbated swelling or other BCRL symptoms afterward.[23,24] These results support the use of resistance training in BCRL management postsurgery and suggests that current restrictive advice regarding loading of the affected arm should be reconsidered. Cormie et al reported similar findings in their study from 2013 where the same patient group exercised at a load of 6 to 8 RM without an increase in swelling.[12] No adverse events appeared in these studies; and thus, they underline the importance of findings in the present validation study.

Limitations

The high ICC values as seen in Table 2 indicate that measurements will be very similar regardless of the rater performing the tests. Although the ICC values are high, some of them have a large CIs, for example, shoulder abduction on both sides and elbow extension. The large CIs could indicate that possible compensatory movements are more difficult for the raters to control for in these subtests—especially shoulder abduction and elbow flexion with the participants in standing and unsupported position. Compensations could be decreased by letting the participants conduct the elbow flexion and shoulder abduction while sitting on a chair against a wall and thereby reducing compensatory movement in the torso. Regarding elbow extension, this exercise is conducted in a triceps machine where strength is being measured bilaterally and where torso position will affect force development; therefore, it can be difficult to ensure that only the affected triceps muscle is developing force. Very precise instruction and correction by raters is required, and deviations will potentially affect the outcome measurements.

Compensatory movements might also be partly responsible for the wide CIs in agreement estimates, as raters have different perceptions of the degree of compensatory movements observed, and different thresholds for when it leads to termination of the test.

With regard to the relative reliability, an acceptable sample size according to de Vet et al[17] has been applied in this study. However, a larger sample size would most likely have contributed to an even better level of the CI for the ICC value.

The modest number of participants is a limitation in this study with regard to the absolute reliability of the test battery. Wide intervals between upper and lower LOA could be due to the small number of participants since the calculation of LOA largely depends on the sample size.[15-17] According to de Vet et al, 50 participants are required to provide a reasonable number of dots in a Bland and Altman plot to estimate the LOA.[16,17]

A possible limitation of this study might be the lack of a training and agreement phase to develop a common standardized agreement of the protocol. However, results of this study showed a high interrater reliability even though the 2 raters had different levels of experience with the dynamic muscle tests; 1 rater received a 3-hour introduction to the tests and practiced testing on pilot participants and had since been testing in the LYCA project. The other rater constructed the test protocol and held the introductory training sessions in connection to starting the LYCA study. In

preparation for this reliability study, they had the same instructions. The results of the current reliability study may provide a realistic picture of testing in clinical practice, where physiotherapists unavoidably will have different levels of experience with maximal muscle testing. Therefore, the results of this study could represent the actual interrater reliability in clinical practice. It can be discussed whether recommendations about preparation before testing were met with regard to developing a common standardized agreement as proposed by Patijn and Remvig.[15,16] They propose testing start with a training and agreement phase prior to the study where the raters develop a common standardized agreement of the protocol.

In this study, only the interrater reliability had been examined, but it would also be relevant to evaluate the intrarater reliability to ensure that the variation when comparing results of repeated measurements by the same rater is as small as possible. This would contribute to the strength of the overall reliability.

Recommendations for Future Research

In clinical practice, it is often a condition that test procedures are conducted by clinicians with different professions and from various settings. The translational value may therefore be increased even more by involving more than 2 raters from different backgrounds, and thereby determining whether satisfactory levels of interrater reliability can be achieved for testers from a wider range of backgrounds. Furthermore, to increase the strength of future research, we recommend that the included raters go through a standardized agreement preparation as proposed by Patijn and Remvig.[15,16]

Conclusion

Results of this study show high relative interrater reliability of dynamic maximal muscle strength tests in women at high risk of developing BCRL, expressed by ICC values higher than 0.80 for all 4 dynamic strength tests individually and 0.99 for the complete test battery. The test battery may therefore be considered a valid clinometric tool to assess muscle strength in women after breast cancer surgery involving ALND. The absolute reliability assessment lacked statistical power but suggested that no systematic bias in testing was present. For higher agreement and consistency to be obtained, an extensive agreement training is likely to be useful, as is a higher number of participants to assess absolute reliability.

Authors' Note

This study originated during the bachelor's degree programme in physiotherapy at University College, Copenhagen, Denmark, 2017

Acknowledgments

The authors wish to thank all the involved women from project LYCA, who extraordinarily participated in this study. We thank Merete Celano Wittenkamp, Charlotte Lanng, Ann Nielsen, and Herlev Hospital for the use of space and equipment outside opening hours. The authors would like to give a special thanks to The Danish Cancer Society Research Center for the collaboration.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Gunn Ammitzbøll https://orcid.org/0000-0003-3227-9475

Supplemental Material

Supplemental material for this article is available online.

References

1. World Health Organization. *Breast Cancer*. Geneva, Switzerland: World Health Organization.
2. World Cancer Research Fund International. *Worldwide Cancer Data*. London, England: World Cancer Research Fund International.
3. Kræftens Bekæmpelse. Statistik om brystkræft. https://www.cancer.dk/brystkraeft-mammacancer/statistik-brystkraeft/. Accessed January 27, 2020.
4. Kræftens Bekæmpelse. Operation for brystkræft. https://www.cancer.dk/brystkraeft-mammacancer/behandling-brystkraeft/operation-brystkraeft/. Accessed January 27, 2020.
5. Lane K, Worsley D, McKenzie D. Exercise and the lymphatic system: implications for breast-cancer survivors. *Sports Med*. 2005;35:461-471.
6. Schmitz KH. Balancing lymphedema risk: exercise versus deconditioning for breast cancer survivors. *Exerc Sport Sci Rev*. 2010;38:17-24.
7. Schmitz KH, Ahmed RL, Troxel AB, et al. Weight lifting for women at risk for breast cancer-related lymphedema: a randomized trial. *JAMA*. 2010;304:2699-2705.
8. Sagen A, Kåresen R, Risberg MA. Physical activity for the affected limb and arm lymphedema after breast cancer surgery. A prospective, randomized controlled trial with two years follow-up. *Acta Oncol*. 2009;48:1102-1110.
9. Kilbreath SL, Refshauge KM, Beith JM, et al. Upper limb progressive resistance training and stretching exercises following surgery for early breast cancer: a randomized controlled trial. *Breast Cancer Res Treat*. 2012;133:667-676.
10. Buchan J, Janda M, Box R, Schmitz K, Hayes S. A randomized trial on the effect of exercise mode on breast cancer-related lymphedema. *Med Sci Sports Exerc*. 2016;48:1866-1874.

11. Courneya KS, Segal RJ, Mackey JR, et al. Effects of aerobic and resistance exercise in breast cancer patients receiving adjuvant chemotherapy: a multicenter randomized controlled trial. *J Clin Oncol*. 2007;25:4396-4404.
12. Cormie P, Galvão DA, Spry N, Newton RU. Neither heavy nor light load resistance exercise acutely exacerbates lymphedema in breast cancer survivor. *Integr Cancer Ther*. 2013;12:423-432.
13. Cormie P, Singh B, Hayes S, et al. Acute inflammatory response to low-, moderate-, and high-load resistance exercise in women with breast cancer-related lymphedema. *Integr Cancer Ther*. 2016;15:308-317.
14. Ammitzbøll G, Lanng C, Kroman N, et al. Progressive strength training to prevent LYmphoedema in the first year after breast CAncer—the LYCA feasibility study. *Acta Oncol*. 2017;56:360-366.
15. Ammitzbøll G, Johansen C, Lanng C, et al. Progressive resistance training to prevent arm lymphedema in the first year after breast cancer surgery: results of a randomized controlled trial. *Cancer*. 2019;125:1683-1692.
16. Patijn J, Remvig L. *Reproducibility and Validity—Protocol Formats for Diagnostic Procedures in Manual/Musculoskeletal Medicine*. Brussels, Belgium: International Academy of Manual/Musculoskeletal Medicine; 2010.
17. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Reliability. In: *Measurement in Medicine—A Practical Guide*. Cambridge, England: Cambridge University Press; 2011:96-149.
18. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2000.
19. Giavarina D. Understanding Bland Altmann analysis. *Biochem Med (Zagreb)*. 2015;25:141-151.
20. Beyer N, Magnusson P, Thorborg K. *Målemetoder I Forebyggelse, Behandling Og Rehabilitering*. 2nd ed. Copenhagen, Denmark: Munksgaard; 2012.
21. Lindahl M, Juul C. *Den Sundhedsvidenskabelige Opgave—Vejledning Og Værktøjskasse*. Copenhagen, Denmark: Munksgaard; 2013.
22. Kræftens Bekæmpelse. Mulige bivirkninger ved kemoterapi. https://www.cancer.dk/hjaelp-viden/kraeftbehandling/behandlingsformer/kemoterapi/bivirkninger-kemoterapi/. Accessed January 27, 2020.
23. Heyward VH, Gibson AL. *Advanced Fitness Assessment and Exercise Prescription*. 7th ed. Champaign, IL: Human Kinetics; 2014.
24. Bloomquist K, Oturai P, Steele ML, et al. Heavy-load lifting: acute response in breast cancer survivors at risk for lymphedema. *Med Sci Sports Exerc*. 2018;50:187-195.