

# The Stepped-Wedge Trial Design: Paving the Way for Cancer Care Delivery Research

YA-HUEI LI, PhD, ELIZABETH MULLETTE, RN, MSN, and  
JEANNINE M. BRANT, PhD, APRN, AOCN®, FAAN

From Billings Clinic Cancer Center,  
Billings, Montana

Authors' disclosures of conflicts of interest are  
found at the end of this article.

Correspondence to: Jeannine M. Brant, PhD,  
APRN, AOCN®, FAAN, Billings Clinic Cancer Center,  
2825 8th Ave N, Billings, MT 59107.  
E-mail: jbrant@billingsclinic.org

<https://doi.org/10.6004/jadpro.2018.9.73>

© 2018 Harborside™

## Abstract

Cancer care delivery research (CCDR) is an emerging field that investigates ways to optimally provide care for patients within complex health-care systems. Novel research designs are essential to efficiently study CCDR research questions. A stepped-wedge trial (SWT) is one such pragmatic design and is similar to a parallel randomized controlled trial (RCT). An SWT design has several advantages. It can examine the clinical effectiveness of an intervention by using participants as the control group, address potential ethical issues, and extend time for trial implementation or policy changes with fewer resources than are used to conduct several RCTs. All participants eventually receive the intervention, which can make the trial more desirable for patient participation. This article aims to introduce and discuss the SWT study design and to encourage future application for CCDR and other oncology-related research.

Cancer care delivery research (CCDR) is defined as “the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and health-care provider and patient behaviors affect access to cancer care, the quality and cost of cancer care, and ultimately the health and well-being of cancer patients and survivors” (National Cancer Institute, 2010).

It is an emerging field and provides hope to build evidence to support practice changes across the country (Kent et al., 2015). Unfortunately, the science is not developing at the rate originally anticipated. A strategic framework is necessary to drive this science forward, along with the use of implementation science models and change theories (Mitchell & Chambers, 2017). Additionally, novel research methods and designs are important to facilitate the research to ensure trials are employed effi-

ciently and successfully. This article will discuss the stepped-wedge trial (SWT) design as one novel approach to CCDR and other oncology-related studies.

## RANDOMIZED CONTROLLED TRIAL DESIGNS

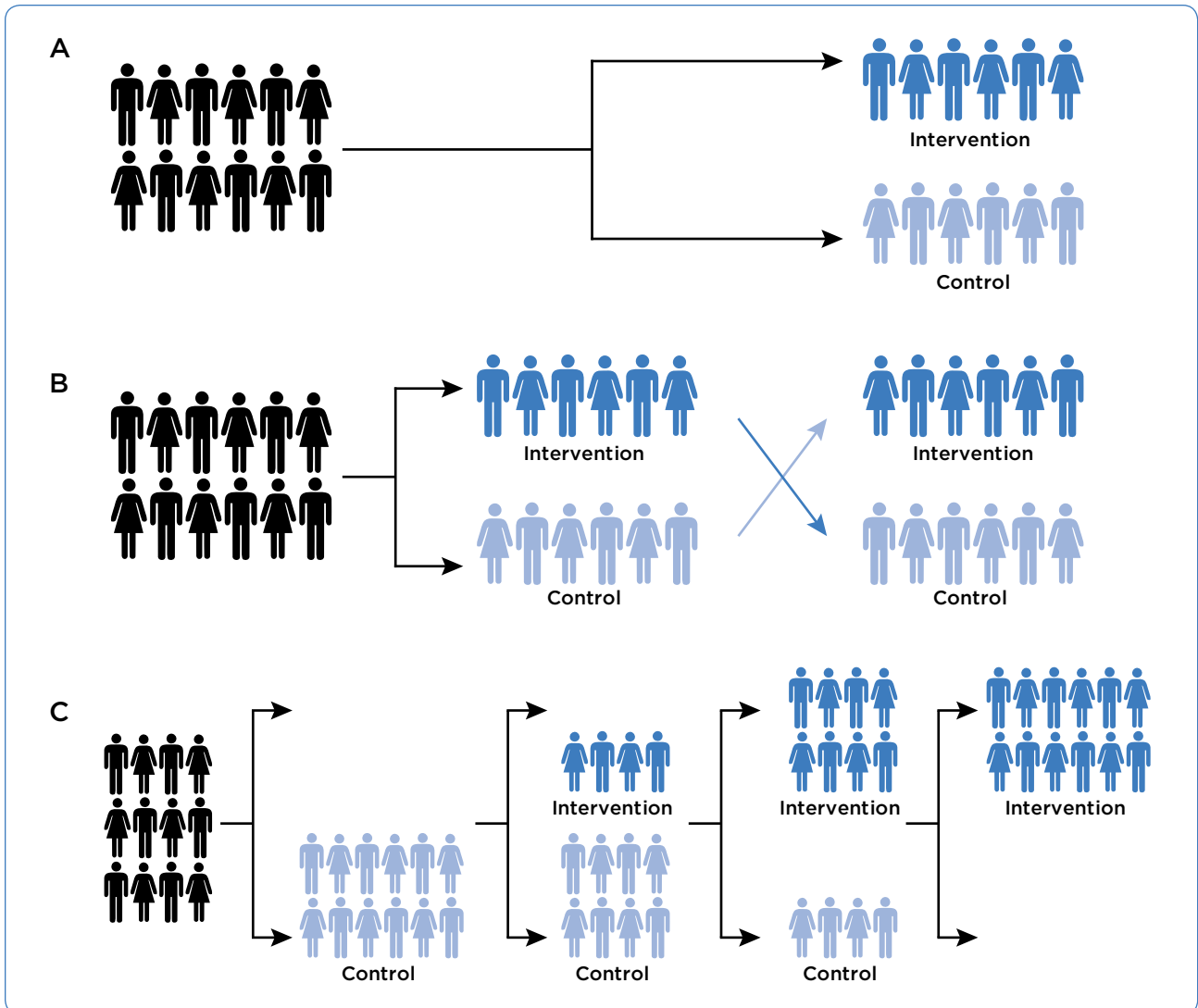
Randomized controlled trials (RCTs) are considered the gold standard for testing clinical effectiveness (Pocock, 1983). In an RCT, all participants (either individually or as part of cluster, e.g., a specific clinic) are randomly assigned to either the control or intervention group. For example, a current CCDR trial, Trial Assessing CSF Prescribing Effectiveness and Risk (TrACER), is investigating the use of a guideline-based colony-stimulating factor (CSF) standing order intervention for neutropenia prophylaxis. Sites (cancer care clinics) were assigned to one of four groups or clusters: (1) usual care and an automated system is already in place; (2) usual care and an automated system is not in place; (3) clinics to have an automated system installed and suggestions for CSF administration; and (4) clinics to have an automated system installed and suggestions for CSF administration along with suggestions for when not to use a CSF. Randomization of clinics prevents contamination between the clusters.

The SWT is a type of RCT using a variation of a crossover design (Figure 1; Cook & Campbell, 1979; Hussey & Hughes, 2007). Unique to the SWT design, all participants are assigned to the control group first and then over time are moved to the intervention group. The design of an SWT is like the crossover study design in that participants switch from one group to another. However, unlike the crossover study design, all participants start out in the control group at the beginning of the study and finish in the intervention group at the end of the study. The order of allocation to the intervention group is randomly assigned as a parallel RCT. At the end of the allocation, all participants will have had the intervention. The SWT is usually preceded by a pilot study that establishes some benefit for participants and deems it ready to move forward to the actual SWT. It can be a useful design for CCDR in that all clinics may desire to receive the intervention as it is perceived beneficial, and yet a control group is needed to fully test the intervention.

## ADVANTAGES AND LIMITATIONS OF THE STEPPED-WEDGE TRIAL DESIGN

The SWT design is a pragmatic RCT design when compared to parallel RCT in that the SWT addresses ethical considerations and logistical limitations (Barker, McElduff, D'Este, & Campbell, 2016; Hemming, Haines, Chilton, Girling, & Lilford, 2015). It resolves the ethical dilemma of a parallel RCT, where the intervention is believed to benefit patients, but the control group cannot receive the intervention due to the design of the study. Because all participants will receive the intervention eventually, it allows individuals to be compared to themselves before and after the intervention, thus controlling for unmeasured confounders. Alternatively, several parallel RCTs would need to be conducted to allow all participants to receive the intervention, but this is often unrealistic due to cost and time constraints. The SWT design also tolerates more heterogeneity in the study. Participants who begin the intervention at the same time form a cluster. Rather than individual participants, clusters switch at each step randomly. Although it takes longer to finish a SWT, it is more cost- and time-efficient than performing several parallel RCTs.

The SWT design does have some disadvantages. First, it is more difficult to design and implement than a parallel RCT. For example, randomizing participants or clusters at each time point, assigning numbers and length of steps, and ensuring that all important elements of the study design are employed requires additional time and effort on the part of the investigator(s). The trial may also take longer to complete, as all participants cannot enter the trial at once and all clusters must finish the intervention. It also involves heavier data collection than other studies and more burden on the researchers and participants, since outcomes are measured at each time point and as participants are entering the trial at different times. Hence, data collection is more complicated. Contamination between participants is another risk; however, this can be controlled by randomizing sites to the steps rather than patients. This would mean that patients at one clinic location are unlikely to talk to patients at another location, thus preventing contamination. Due to the study design, the sample size calculation is inherently more difficult. Finally, the SWT design is not feasible for multiple treatment options.



**Figure 1.** Treatment schedules for randomized controlled trial designs. (A) Randomized controlled trial: Participants are randomly assigned into control and intervention groups by investigators. (B) Crossover: A total of two sections are included in the study. Participants are assigned to either a control or intervention group in section 1. After taking the measurements at the end of the first session, the participants are switched to another arm (the control group becomes the intervention group; the intervention group becomes the control group). At the end of the second session, comparison measurements are taken again from both groups. (C) Stepped-wedge design: All participants experience being in the control and the intervention groups. At the beginning of the trial, all participants are assigned to the control group. At a later point in time (time length determined by design), a certain proportion of the participants are rolled into the intervention group sequentially over time. The order of enrollment into the intervention group is random. By the end of the random allocation, all participants will have received the intervention.

### A STEPPED-WEDGE TRIAL STUDY EXAMPLE: APPLICATION IN ONCOLOGY

Caminiti, Iezzi, and Passalacqua (2017) published a study, “Effectiveness of the HuCare Quality Improvement Strategy on health-related quality of life in patients with cancer: Study protocol of a

stepped-wedge cluster randomised controlled trial (HuCare2 study),” which is an example of an SWT used in CCDR. Investigators sought to examine how the HuCare Quality Improvement Strategy (HQIS) improved health-related quality of life in patients with cancer (Caminiti et al., 2017). The HQIS, comprised of integrating six interna-

tionally recommended psychosocial interventions into practice, was initially developed and proved feasible in 28 medical cancer centers across Italy (Passalacqua et al., 2016).

A randomized controlled trial was recommended as the next step to measure the success of implementing these evidence-based interventions compared to standard care. Specifically, investigators aimed to examine whether the HQIS improved health-related quality of life. The primary endpoint was the mean change of social and emotional scores between baseline (prior to the intervention) and 3 months after the intervention. A 3-month follow-up was chosen because a peak of emotional distress was observed in patients after treatment initiation. Secondary outcomes of interest included the impact of HQIS on quality-of-life domains in the long term (6 months and 1 year) according to specific patient characteristics (e.g., anxiety, depression, cancer types) and clinician adherence to the HQIS intervention.

The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 was used to assess the global quality of life (Apolone, Filiberti, Cifani, Ruggiata, & Mosconi, 1998), and the Hospital Anxiety and Depression Scale was applied to evaluate patients' psychological conditions (Costantini et al., 1999). Unmet social needs of patients were assessed using the Needs Evaluation Questionnaire (Tamburini et al., 2000). To assess the quality of the inter-

vention, investigators calculated the percentage of clinical staff who completed the HQIS training and applied patient communication skills per the trained intervention along with the percentage of patients who received recommendations from the intervention.

Fifteen centers agreed to participate in the study. Five centers were then grouped together into a cluster, for a total of three clusters. Each cluster was randomly assigned to the HQIS intervention with the sequence generated by SAS, a statistical analysis software. The study period was divided into five time periods (epochs), each lasting 4 months. The first epoch included all centers in the control group, whereas the last epoch included all centers in follow-up. The intervention was delivered sequentially during epochs 2 to 4. Figure 2 illustrates the stepped-wedge cluster randomized controlled study design used in HuCare2.

To determine the number of patients in each cluster, investigators anticipated a total of 720 cancer patients enrolled in the study, with approximately a 20% dropout rate at follow-up. Based on a power of 80% and a two-tailed alpha of 5%, they planned for 60 patients in each cluster. For example, a total of 180 patients were in the control group at epoch 1, whereas 60 patients were still in the control group and 60 patients were in the postintervention group in epoch 3. To satisfy the 3-month follow-up requirement for all clusters, it took a total of 20 months to finish this study. The

	Epoch 1 (Months 1-4)	Epoch 2 (Months 5-8)	Epoch 3 (Months 9-12)	Epoch 4 (Months 13-16)	Epoch 5 (Months 17-20)	Patient numbers
Cluster 1	Control	HQIS intervention	Postintervention	Postintervention	Postintervention	<b>240/4 clusters</b>
Cluster 2	Control	Control	HQIS intervention	Postintervention	Postintervention	<b>240/4 clusters</b>
Cluster 3	Control	Control	Control	HQIS intervention	Postintervention	<b>240/4 clusters</b>
<b>Patient numbers</b>	<b>180/3 clusters</b>	<b>120/2 clusters</b>	<b>120/2 clusters</b>	<b>120/2 clusters</b>	<b>180/3 clusters</b>	<b>Total: 720/12 clusters</b>

**Figure 2.** Stepped-wedge cluster randomized controlled trial study design in the HuCare2 study. A cluster is the unit of the trial randomly selected to the HQIS intervention. Each cluster includes five centers located in the same region. A total of 60 patients are calculated for each cluster, with a 20% dropout rate at follow-up and 80% power and two-tailed alpha of 5%. In epochs 1 and 5, there will be 180 patients in either the control or postintervention group. In each progressive time sequence, an increasing number of patients in the postintervention group is observed. Epoch 2: 120 patients in the control group; epoch 3: 60 patients in the control group, and 60 patients in the postintervention group; epoch 4: 120 patients in the post-intervention group. HQIS = HuCare Quality Improvement Strategy. Information from Caminiti et al. (2017).

study is currently underway, and no results are presented in the article. However, the study design provides an exemplar for an SWT that can be used in cancer care delivery and other types of oncology-related research.

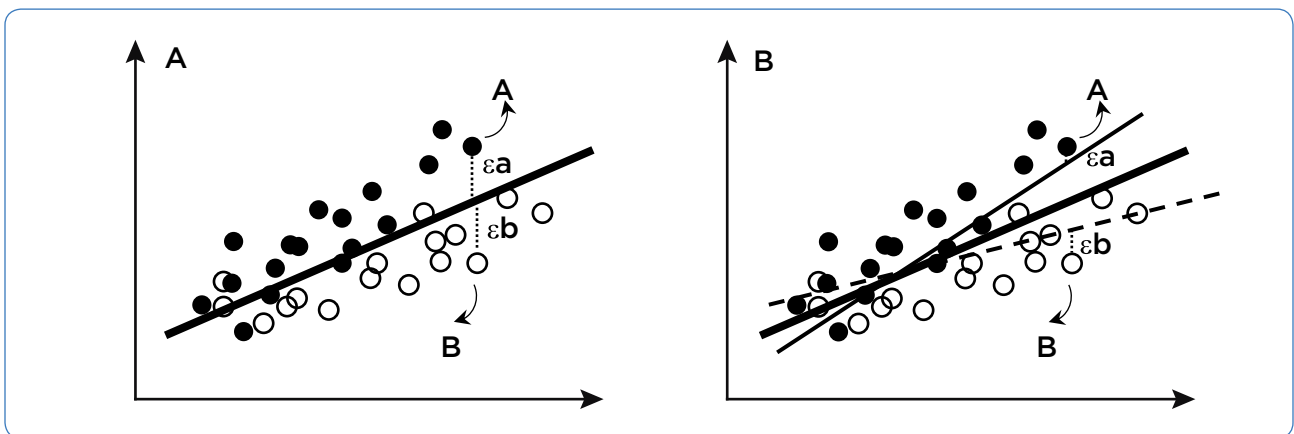
## STATISTICAL METHODS

A variety of statistical methods can be used to analyze study outcomes from the SWT design. For the study example, investigators plan to use a beta-binomial regression model to determine predictors (e.g., anxiety, depression, cancer types) of the intervention's success. The model is considered binomial because the outcome variable is a two-level response (improved quality of life or not) to the intervention. Other advanced analytical procedures commonly employed include a generalized estimating equation (GEE) and generalized linear mixed model (GLMM; Barker et al., 2016). Figure 3 compares the GEE and GLMM.

A GEE is a type of generalized linear model (GLM) that tries to quantify how much each predictor influences the outcome. For example, maybe the clinic setting alone has an influence or maybe it could be the individual delivering the intervention and their background training. Patient demographics such as age, sex, anxiety, and edu-

cational level are all characteristics that can influence an outcome. Responses from the individuals are considered the actual data, but this can be very different when data are merged together to create a model. The GEE takes these data, recognizes the error that exists with the data, but estimates the error based on the entire population and does not consider the cluster population. This can lead to large error terms, as individuals may vary substantially from the general population.

The GLMM offers a better approach and may be referred to as a mixed, hierarchical, or multi-level model. The GLMM estimates both fixed and random effects within a study at the same time. To further explain this, fixed effects represent the data from the entire population. For example, in the Caminiti study, 720 patients represent the entire population for the fixed effect. Their data will be modeled using an intercept and slope for the entire sample. And yet the 5 clinics participating in the study can each be represented by 4 clusters of 60 patients each. Each of these clusters will be modeled again with a group intercept and slope. Their data, which is represented on a line, will look slightly different from the overall population data of 720 patients. The difference between each of the clusters and the total population estimate



**Figure 3.** Schematic diagram of the generalized estimating equation and the generalized linear mixed model. (A) A generalized estimating equation allows researchers to draw a line (black bold line) between all black and white dots, representing the population means. The dot A and dot B are two participants among the population in the trial. The difference between the population mean and the individual is considered the error term. For example, the error term of dot A is  $\epsilon_a$ , while the error term of dot B is  $\epsilon_b$ . (B) A generalized linear mixed model allows researchers to examine the fixed (population mean) and random effects (difference between the population mean and subpopulation mean). As a result, two lines are depicted: a thin black line and a dotted line, representing two subpopulation (black dots and white dots) means. After taking the fixed and random effects into consideration, not only are the estimates accurately estimated among the clusters, but the error terms are also reduced ( $\epsilon_a$  and  $\epsilon_b$  are smaller in B than in A).

is the random effect, which is also considered in the overall model. The error term is an estimate of how far the individual is from the group's clustered data. If the participants are homogeneous, the random effects will be close to zero. The GLMM is considered a better option than the GEE when there are greater numbers of clusters in the trial with a lower likelihood of a type I error, which is finding a difference when it really is not there (Tabachnick & Fidell, 2013).

### IMPLICATIONS FOR STEPPED-WEDGE TRIAL USE IN CANCER CARE DELIVERY RESEARCH

The SWT is one type of design that can facilitate CCDR. Caminiti and colleagues (2017) illustrated the advantages and provided rationale as to why they chose this design. First, their goal was to widely disseminate the HQIS intervention at a national level to investigate the changes of cancer patients' quality of life after implementing HQIS. It would have been impossible to train numerous individuals from multiple centers to deliver the intervention simultaneously. Second, the design offered the ability to randomize sites so that contamination would not occur within a center. Site randomization is a common strategy in CCDR and is favored when contamination is a concern (Kent et al., 2015). Finally, the study was ethically acceptable in that all centers would eventually receive the intervention.

Overall, the SWT design is pragmatic for participants, investigators, and policy makers. It can be executed within a hospital, region, nationally, or internationally to globally disseminate an intervention and ensure that all sites receive the intervention. Stepped-wedge trials have the potential to benefit all sites and/or patients that participate in the trial. Most importantly, with limited resources and with the complexity of CCDR interventions, SWT design allows evidence-based interventions to be implemented widely in a sequential manner, not to overwhelm investigators, but rather to impact patient outcomes globally. ●

#### Disclosure

Dr. Brant has served on the speakers bureau for Genentech. Ms. Li and Ms. Mullette have no conflicts of interest to disclose.

#### References

Apolone, G., Filiberti, A., Cifani, S., Ruggiata, R., & Mosconi, P. (1998). Evaluation of the EORTC QLQ-C30 questionnaire: A comparison with SF-36 Health Survey in a cohort of Italian long-survival cancer patients. *Annals of Oncology, 9*(5), 549–557.

Barker, D., McElduff, P., D'Este, C., & Campbell, M. J. (2016). Stepped wedge cluster randomised trials: A review of the statistical methodology used and available. *BMC Medical Research Methodology, 16*, 69. <https://doi.org/10.1186/s12874-016-0176-5>

Caminiti, C., Iezzi, E., & Passalacqua, R. (2017). Effectiveness of the HuCare Quality Improvement Strategy on health-related quality of life in patients with cancer: Study protocol of a stepped-wedge cluster randomised controlled trial (HuCare2 study). *BMJ Open, 7*(10), e016347. <https://doi.org/10.1136/bmjopen-2017-016347>.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, IL: Rand McNally.

Costantini, M., Musso, M., Viterbori, P., Bonci, F., Del Mastro, L., Garrone, O.,...Morasso, G. (1999). Detecting psychological distress in cancer patients: Validity of the Italian version of the Hospital Anxiety and Depression Scale. *Supportive Care Cancer, 7*(3), 121–127.

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting. *BMJ, 350*, h391. <https://doi.org/10.1136/bmj.h391>

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials, 28*(2), 182–191. <https://doi.org/10.1016/j.cct.2006.05.007>

Kent, E. E., Mitchell, S. A., Castro, K. M., DeWalt, D. A., Kaluzny, A. D., Hautala, J. A.,...Clauser, S. B. (2015). Cancer care delivery research: Building the evidence base to support practice change in community oncology. *Journal of Clinical Oncology, 33*(24), 2705–2711. <https://doi.org/10.1200/JCO.2014.60.6210>

Mitchell, S. A., & Chambers, D. A. (2017). Leveraging implementation science to improve cancer care delivery and patient outcomes. *Journal of Oncology Practice, 13*(8), 523–529. <https://doi.org/10.1200/JOP.2017.024729>

National Cancer Institute. (2010). NCI Community Oncology Research Program: Research areas. Retrieved from <http://ncorp.cancer.gov>

Passalacqua, R., Annunziata, M. A., Borreani, C., Diodati, F., Isa, L., Saleri, J.,...Caminiti, C. (2016). Feasibility of a quality improvement strategy integrating psychosocial care into 28 medical cancer centers (HuCare project). *Supportive Care Cancer, 24*(1), 147–155. <https://doi.org/10.1007/s00520-015-2756-z>

Pocock, S. J. (1983). *Clinical trials: A practical approach*. Hoboken, NJ: John Wiley & Sons Ltd.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6<sup>th</sup> edition). Boston, MA: Pearson.

Tamburini, M., Gangeri, L., Brunelli, C., Beltrami, E., Boeri, P., Borreani, C.,...Trimigno, P. (2000). Assessment of hospitalised cancer patients' needs by the Needs Evaluation Questionnaire. *Annals of Oncology, 11*(1), 31–37.