Evaluation and analysis of stepped wedge designs
Zhan, Zhuozhao

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
STRENGTHS AND WEAKNESSES OF A STEPPED WEDGE CLUSTER RANDOMIZED DESIGN: ITS APPLICATION IN A COLORECTAL CANCER FOLLOW-UP STUDY

Z. Zhan
E. R. van den Heuvel
P. M. Doornbos
H. Burger
C. J. Verberne
T. Wiggers
G. H. de Bock

This chapter has been published in Journal of clinical epidemiology, Volume 67, Issue 4, (2014) [26].
ABSTRACT

Objectives: To determine the advantages and disadvantages of a stepped wedge design for a specific clinical application.

Study design and settings: The clinical application was a pragmatic cluster randomized surgical trial intending to find an increased percentage of curable recurrences in patients in follow-up after colorectal cancer. Advantages and disadvantages of the stepped wedge design were evaluated, and for this application, new advantages and disadvantages were presented.

Results: A main advantage of the stepped wedge design was that the intervention rolls out to all participants, motivating patients and doctors, and a large number of patients who were included in this study. The stepped wedge design increased the complexity of the data analysis, and there were concerns regarding the informed consent procedure. The repeated measurements may bring burden to patients in terms of quality of life, satisfaction, and costs.

Conclusions: The stepped wedge design is a strong alternative for pragmatic cluster randomized trials. The known advantages hold, whereas most of the disadvantages were not applicable to this application. The main advantage was that we were able to include a large number of patients. Main disadvantages were that the informed consent procedure can be problematic and that the analysis of the data can be complex.
2.1. Introduction

Back in 1967, the concept of a pragmatic trial was proposed by Schwartz and Lellouch.[19] A pragmatic trial is designed to evaluate the effectiveness of interventions in real-life routine practice conditions [16]. Its aim is to answer the question “Does the intervention work when used in real-life practice condition?” Thus, in a pragmatic trial, there are no or minimal exclusion criteria required. It also implies that pragmatic trials are normally used when there is a priori knowledge of the efficacy of the intervention under study. In addition, a pragmatic trial is often concerned with complex interventions, for example, routine screening of the disease rather than a pharmaceutical intervention, and it is typically compared with care as usual.

To answer the question whether the intervention works when used in real-life practice, it is very common to apply a cluster randomized design. One of the reasons for choosing a cluster design is the concern of contamination.[23] Cluster randomization may reduce the risk that the intervention under study is unintentionally mixed up with care as usual, the intervention of the control group.[1, 18, 22] Another reason is that the intervention can be performed more easily in clusters as a large number of participants can make it impractical to introduce a new treatment on an individual level, for example, when medical resources are low or when there are costly expenses.[6]

The stepped wedge design is a unique design suitable to answer the question whether the intervention works when used in real-life practice.[3, 10] This design allows for a controlled stepwise introduction of an intervention to a population.[3, 10] Although not per definition, the stepped wedge design is a design that is mostly performed as a cluster design.[15] In this design, participants start in the control group, and at predefined time points, a cluster of participants are switched to the intervention
group in a random order (known as “steps”). From the moment of switching until the end of the study, they will stay in the intervention group.[10]

When the stepped wedge design is compared with other designs, there are several advantages and disadvantages of choosing such a design. The aim of the present article was to determine the advantages and disadvantages of a stepped wedge cluster randomized design for a specific clinical application. An overview of the literature regarding the advantages and disadvantages of the stepped wedge design will be given. The clinical application is the colorectal cancer (CRC) follow-up study (CEAwatch, Netherlands Trial Register 2182). Based on a point-by-point evaluation, it was analyzed how these advantages and disadvantages apply to our specific clinical trial. Furthermore, some new trial-specific advantages and disadvantages of the stepped wedge design in this application, which were not mentioned in the literature, will be added.

2.2. METHODS/DESIGN

2.2.1. ADVANTAGES AND DISADVANTAGES OF STEPPED WEDGE DESIGNS

Under the circumstance when there is prior evidence that the intervention under study will do more good than harm, rather than clinical equipoise, it is considered not ethical to withhold or withdraw an intervention from participants.[3] As the stepped wedge design provides unidirectional sequential rollout of the intervention, all participants will get the intervention during the study. Additionally, the stepped wedge design can be a good option in trials in which it is not possible to introduce the intervention to all participants at once because of logistic, financial, or practical reasons as the design introduces the intervention over multiple moments.[3] The stepped wedge design is considered to be a strong design to evaluate effects on a population level.[9] It is favored over some other trial designs because it provides an opportunity to measure possi-
ble effects of the time of the intervention and to investigate the effects of underlying temporal changes.[3] The stepped wedge design is more efficient than others as it may reduce the required number of clusters compared with other classic cluster designs.[10, 25] Although between-cluster variation affects the statistical power in a parallel clustered randomized design, the power appears to be relatively insensitive to between-cluster variation in the stepped wedge design.[10] Stepped wedge design requires fewer clusters because the power of the design is mainly determined by within-cluster variations.[10]

One of the drawbacks of the stepped wedge design is that it takes a longer time to perform.[13] Because of the nature of a stepwise introduction of the intervention, the trial duration of a stepped wedge design will be the duration of a classic cluster randomized trial multiplied by the number of steps. Especially for clusters that started later have to wait longer depending on the duration of each step, it may cause them to switch into interventions or dropping out. This will then increase the risk of attrition. In addition, the repeated measurements of the stepped wedge design put a heavy burden on patients, caregivers, and researchers.[13] Another concern is that the stepped wedge design may increase the risk of contamination in a cluster, especially when the intervention is believed to be superior to control.[13] It is also very hard to use blinding because both patients and assessors are aware of the step switch.[3] From a statistical perspective, there are also some disadvantages of using the stepped wedge design. Mentioned by Hussey [10], a delay in the treatment effect reduces the power of the design. Moreover, the analysis of the stepped wedge design is more complex.[15]

For a summary of the literature on the advantages and disadvantages of the stepped wedge design, see Table 2.2, first column. Based on a point-by-point evaluation, it was analyzed how these advantages and disadvantages applied to the specific clinical trial CEAwatch (Netherlands
Furthermore, some application-specific advantages and disadvantages of the stepped wedge design, which has not been mentioned in literature, were added.

2.2.2. Clinical application to the CRC follow-up (CEAwatch)

The tumor marker carcinoembryonic antigen (CEA) has long been known to be important in signaling recurrent disease in CRC.[20] Intensive follow-up schedules including CEA measurements are correlated with better survival rate than schedules not using CEA measurements [11], and serial measurements of CEA are recommended for use in CRC follow-up.[4, 14] Other studies also confirmed a reduction of mortality rate and an improvement in curative reoperation rate with intensive surveillance.[21] In a phase 2 trial, monthly CEA measurements were done with a threshold of two consecutive rises of more than 10%.[7] The trial showed both high sensitivity and specificity for detection of recurrences using serial CEA rises rather than absolute values. Given this evidence, in CEAwatch, a new intensified follow-up scheme including frequent CEA measurements and CEA-triggered imaging in detecting recurrent disease with curative possibilities in CRC patients was compared with care as usual.

Patients

Patients with American Joint Committee on Cancer stage I, II, and III CRC after R0 resection, who were surgically operated, were eligible. Patients who received adjuvant chemotherapy were eligible after termination of adjuvant therapy. Patients who were not medically fit for metastasectomy, patients diagnosed with other malignancies (except skin basocellular carcinoma), and patients with metachronous metastases at the start of the study were excluded.
2.2. Methods/design

The follow-up care as usual
The control or “care-as-usual” follow-up consisted of follow-up as recommended in the national guideline in the Netherlands (www.tinyurl.com/coloncarcinoma) including an outpatient clinic visit every 6 months for the first 3 years and an annual visit in years 4 and 5. Liver ultrasound and chest x-ray were recommended at each clinic visit. CEA was measured every 3–6 months in the first 3 years and each year in the last 2 years. No monitoring of compliance with this recommendation was provided.

The CEAWatch follow-up
The intensified follow-up protocol adhered to bimonthly CEA measurements and yearly imaging in the first 3 years and trimonthly CEA measurements in the fourth and fifth years of follow-up (Fig. 2.1). Outpatient clinic visits with imaging of chest and abdomen were performed annually in the first 3 years. The threshold value used was a 20% rise compared with the latest CEA value, followed by a threshold of any rise respect to the last measurement after 1 month. In case of two consecutive rises in CEA, a computed tomographic (CT) scan of chest and abdomen was advised for localization of potential metastatic disease. The coordination of this process was supported by an automatic computer system.[20, 24] Doctors were given an alert when a CT scan was indicated because of a consecutive rise in CEA or when patients forget to go for a CEA assessment. CEA values were communicated to the patients by an automatically generated letter, including a laboratory form for the next CEA measurement.

Study design
The hospitals were randomly grouped into five clusters that were changed from the usual follow-up schedule to the intensive follow-up schedule at different time points. Cluster crossover from the control schedule to the intervention schedule occurred in one direction only and once every 3 months (Table 2.1). Randomization of the crossover moments of
Figure 2.1 | CEA-Watch follow-up and the care-as-usual follow-up. *Local differences and adjustment by individual hospitals allowed.

the clusters was performed independently by Trial Coordination Center Groningen (www.tcc.umcg.nl). CEAWatch was approved by the Medical Ethics Committee of the University Medical Center Groningen and the local ethics committees of all participating centers. For an overview of the procedures, see Fig. 2.2.

Main Outcome
The primary outcome measures were the proportion of resectable recurrences among all recurrences and the time to and probability of detection of recurrent disease in the intervention protocol compared with the control protocol.

Data Collection
In the participating hospitals, the eligible patients were identified using the diagnosis or operation code(s). At the end of the study, this search
2.2. Methods/design

Table 2.1 | Progression of control (0) and intervention group (1) over time periods (t) in CEA-Watch study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>721</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>456</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>613</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>630</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>803</td>
</tr>
<tr>
<td>Number of participants</td>
<td>2,498</td>
<td>2,484</td>
<td>2,503</td>
<td>2,409</td>
<td>2,255</td>
<td>1,946</td>
<td>3,223</td>
</tr>
</tbody>
</table>

was validated against the database of the Dutch Comprehensive Cancer Center. In this database, all newly diagnosed malignancies are registered based on the automated pathologic archive. After all eligible patients were identified, patient and tumor characteristics were exported from the Dutch Surgical Colorectal Audit (DSCA) into a password-protected database. DSCA is an obligatory national data bank that gathers all relevant information on surgically treated CRC patients, allowing a valid registration of all CRC patients in the Netherlands, without any missing baseline characteristics (www.clinicalaudit.nl/dsca). Per hospital, there was one study coordinator. The study coordinators were uniformly trained to identify new eligible patients, inform patients about the study, and collect the follow-up data. The study coordinators were continuously monitored by one of the investigators.

Power calculation

The expected percentage of resectable recurrences was 10% in the control protocol and 25% in the intensified protocol.[2, 17] Given a significance level of 5% and a power of 80%, 115 patients with recurrent disease in both groups were needed. Given an expected recurrence rate of 25% [12],
Figure 2.2 | Study procedures in CEA-Watch.

460 patients per group were needed. Given the cluster randomization, we assumed a correlation of 0.1 between hospitals, yielding a correction factor of 1.71. Therefore, a minimum of about 800 patients per group was needed.

**DATA ANALYSIS**

To compare the effect of the intensified follow-up with the control follow-up protocol regarding the proportion of resectable recurrences, a conditional logistic regression analysis, with hospital as the stratification variable, was performed. Cox proportional hazards model formed the basis of the analysis of the time-to-event data (recurrence or curable recurrence). Hereby, the follow-up protocols were used as a time-dependent variable because the switch time in follow-up was dynamic for patients. The time from the operation until the participation in the study created
left truncated data, and to correct for this, a delayed entry variable was implemented. Again, the analysis was stratified by hospital.

**CURRENT STATUS OF CEAwatch**

Inclusion of patients \( n = 3223 \) started from October 2010 and ended in July 2012. Every 3 months, there was a switch from the care-as-usual follow-up to CEAwatch follow-up, which successfully took place. The first switch was in January 2011. The results from this study will be published separately.

### 2.3. RESULTS

#### 2.3.1. ADVANTAGES

The logistic difficulties of implementing the intervention everywhere at once was one of the considerations to choose for a stepped wedge design for the CEAwatch study (Table 2.2).[3] To start the study in 11 hospitals, approvals from 11 local administration institutes were needed. In each participating hospital, the eligible patients had to be identified, patient and tumor characteristics had to be extracted, and study coordinators had to be trained to be able to identify new eligible patients, inform patients about the study, and collect the follow-up data before the study could start. Besides that, the automatic computer system that was used to support the implementation of the intervention under study had to be adapted to the local hospital system before it could be implemented in the hospitals. Thus, it was not feasible to implement the intervention to different hospitals simultaneously.

The stepped wedge design is considered more ethical than other (classic) cluster randomized controlled designs when the intervention is believed to do more good than harm.[3] In our case, there was enough evidence to support the follow-up of patients with CRC with frequent
### Table 2.2 | Advantages and disadvantages of stepped wedge design and their application to CEAwatch

<table>
<thead>
<tr>
<th>General</th>
<th>Application of general advantages/disadvantages to CEAwatch</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td>CEAwatch involved 11 hospitals all around the country, it was very hard to implement the intervention simultaneously, and the specialized software used in this trial also took time to be adjusted hospital by hospital. The stepped wedge design provides an opportunity to prepare for the implementation in the control period during the trial.</td>
</tr>
<tr>
<td>Good alternative when interventions cannot be implemented to all clusters simultaneously because of practical, logistic, or financial constraints.[3]</td>
<td>The CEA measurements and the intensive follow-up protocol were proven effective from individual-level trials. This point of advantage was motivation for patients and doctors to participate. This helped to increase the size of the sample and consequentially increased the power of the trial.</td>
</tr>
<tr>
<td>If there is a prior belief that the intervention will do more good than harm (rather than clinical equipoise), it is considered not ethical to withhold/withdraw intervention from participants. The stepped-wedge design provides sequential rollout of the intervention for all participants.[3]</td>
<td>The end points were time to event and events, which need a certain time period before they are observed. In combination with the relative short period (3 months) between switches, the underlying temporal changes could not be appropriately investigated. Because of the longitudinal setup of the CEAwatch, this point is not applicable for the study.</td>
</tr>
<tr>
<td>Provide opportunity to measure possible effects of time of intervention and to investigate the effects of underlying temporal changes.[3]</td>
<td>CEAwatch intended to include as many hospitals as possible and has no issue regarding the number of clusters. Thus, the study did not make use of this benefit.</td>
</tr>
<tr>
<td>Reduces the number of clusters.[10]</td>
<td>Thus far, this claim was proven under circumstances of very simple settings of a stepped wedge design. It is not sure whether a stepped wedge design will be a more efficient design than other designs in our more complex setting (left truncation, dynamic, or time-varying intervention and stratification in the Cox proportional hazards approach).</td>
</tr>
<tr>
<td>It is more efficient than other cluster randomized controlled trial design.[25]</td>
<td></td>
</tr>
</tbody>
</table>
## 2.3. RESULTS

<table>
<thead>
<tr>
<th>General</th>
<th>Application of general advantages/disadvantages to CEAwatch</th>
</tr>
</thead>
</table>

### Disadvantages

<table>
<thead>
<tr>
<th>劣势</th>
<th>优势</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General</strong></td>
<td>优势</td>
</tr>
<tr>
<td>长期试验持续时间和增加的脱落风险。[13]</td>
<td>因为要求的纵向设置用于CEAwatch试验（多次访问），使用分层楔式设计并不延长试验持续时间与其他设计相比。可选方案持续时间用于随访有5年。使用分层楔式设计进行这项研究时，如果要调查加强随访的效率，就需要有类似持续时间（例如3-5年），在设计中调查这一点是必要的。</td>
</tr>
<tr>
<td>反复测量对患者、照顾者和研究人员造成沉重负担。[13]</td>
<td>这种集群随机试验中，并非所有患者都必须通过加强的随访。当加强方法没有预期那样有效时，这将是有益的。患者是否认为CEAwatch 加强的随访是一个更高的负担，这一点至关重要，所以我们希望用质量生命和成本效益研究进行测量。</td>
</tr>
<tr>
<td>增加的污染风险。[13]</td>
<td>污染风险有限，因为实施了定制软件来支持加强的CEA 随访。</td>
</tr>
<tr>
<td>处理效果的延迟降低了设计的功率。[10]</td>
<td>这是真实的，但是考虑到从较大的样本中获得的益处，由于处理效果的延迟导致的功率降低并不大。</td>
</tr>
<tr>
<td>缺乏盲法</td>
<td>信息偏倚也可以考虑作为对患者和医生治疗反应的处理，用于CSEAwatch这样的实践性试验。</td>
</tr>
<tr>
<td>分析设计复杂。[15]</td>
<td>CEAwatch具有相当复杂的网络设计，我们将这一点作为该设计的主要劣势。</td>
</tr>
</tbody>
</table>

### CEAwatch specific

<table>
<thead>
<tr>
<th>劣势</th>
<th>优势</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantage</strong></td>
<td>优势</td>
</tr>
<tr>
<td><strong>Disadvantage</strong></td>
<td>优势</td>
</tr>
<tr>
<td><strong>Advantage</strong></td>
<td>随机招募参与者在CEAwatch期间更容易。这允许医院在控制期中用相同的标准纳入患者，并在研究期间纳入新患者。</td>
</tr>
<tr>
<td><strong>Disadvantage</strong></td>
<td>询问所有患者在基线时签署知情同意书未被医院的伦理委员会批准。这被认为对患者来说是不可接受的。</td>
</tr>
</tbody>
</table>
CEA measurements and CEA-triggered imaging in detecting recurrent disease with curative possibilities. Because of the preferences of surgeons for the intervention under study, this ethical advantage was an important motivation for doctors to participate in the trial. As we had no published evidence that the repeated CEA measures in the intervention under study was not a burden to the patient associated with an increase of costs, data on secondary outcomes such as patient satisfaction, quality of life, and costs were collected.

Another advantage of the stepped wedge design is that this design provides an opportunity to measure possible effects of time of the intervention and investigate the effects of underlying temporal changes because of its longitudinal settings. The CEAwatch study could not really benefit from this point because the longitudinal setting was needed to obtain or collect the events. Besides this, the periods before the switches were relatively short (only 3 months), making it less attractive to model time trends in the analysis of the events. Furthermore, the inclusion of patients was dynamic, complicating such a temporal analysis.

Another general advantage of the stepped wedge design is that it reduces the required number of clusters as the design is relatively insensitive to variations of the intercluster correlation. This might be beneficial for trials that have limited resources and cannot include enough clusters, but for the CEAwatch study, the number of clusters was sufficient and there was no need to include as many hospitals as possible. Thus, this advantage was not one of the considerations to choose for the stepped wedge design. It is claimed that a stepped wedge design is more efficient than other cluster randomized controlled designs. Thus far, this claim was proven under circumstances of very simple settings of a stepped wedge design. It is not sure whether a stepped wedge design was a more efficient design than other designs in our more complex application, which required left truncation, a dynamic or time-varying intervention variable,
and stratification in the Cox proportional hazards approach.

An advantage of the stepped wedge design not mentioned in literature, but very important in our study, is that by the use of the stepped wedge design, we were able to include a large number of patients. In the CEAwatch study, eligible patients were identified before the start of the study, and new patients were included during the time of the study. An advantage of this approach was that patient selection was less vulnerable to selection bias. A second advantage of this approach was that the group of participants consisted of patients who were already in follow-up on the date of the start of the study and those who became eligible during the study period.

2.3.2. Disadvantages

One of the disadvantages of the stepped wedge design is that it takes longer than the more traditional designs.[13] However, because of the longitudinal setup of the CEAwatch study, using a stepped wedge design did not extend the trial duration compared with other designs. As the follow-up for eligible patients was in principle 5 years, it was essential to have comparable trial duration (eg, 3–5 years) to investigate the effectiveness of the intensive follow-up routine no matter what kind of design would have been used. As a consequence, in this case, other designs would not have shortened the trial duration substantially.

Another drawback of the stepped wedge design is the heavy burden on patients, caregivers, and researchers caused by the necessary repeated measurements.[13] As these repeated measurements could not be avoided because of the nature of the intervention in the CEAwatch study, other designs would also have had this problem. On the other hand, in a cluster randomized trial, not all patients have to go through the intensified approach. This could be beneficial when the intensified approach would not be as effective as it was hypothesized. Whether patients also viewed
the CEAwatch intensified follow-up as a burden is of course critical, it was decided to investigate this in the study as secondary outcomes.

Another concern mentioned by Kotz et al. [13] is the increased risk of contamination and attrition. The contamination in a cluster in the CEAwatch study was limited to a minimum because of the automated software system that was used to trigger follow-up schedules. Thus, it would have been very hard that certain patients would still be scheduled under the care-as-usual when the hospital would have changed to the intensified intervention. The attrition problem in the study was mainly due to the long trial duration, which would most likely also have occurred in other types of designs.

When designing a stepped wedge and estimating its sample size, it is suggested by Hussey and Hughes [10] that researchers should take into account the delay in treatment effect as the effect of such delay is a reduction of the power of the design. However, given the nature of the outcome, the delay in treatment effect was considered to be not a problem for the CEAwatch study. It was somewhat compensated with the inclusion of patients who had surgery before the start of the study.

Although it is true that using blinding in the CEAwatch is impossible, it is not typical for a stepped wedge design. In addition, the effects might be not as strong as it is claimed to be.[3] As a pragmatic trial, CEAwatch was interested in studying the responses of patients in a real-life situation. As a consequence, the awareness of the intervention could be accepted as part of the responses to treatment.

Furthermore, it is mentioned that the analysis of the stepped wedge design is complex.[15] This was considered a main disadvantage of the design in CEAwatch. The complexity comes from different sources. One source is the issue with the delayed entry of patients into the study, and another source is the dynamic nature of the inclusion of patients and the switch moments that require a time-varying intervention variable into
the survival analysis. The hospitals were addressed by stratification in the analyses, but they may also be considered as random, which would be typical in the more classical cluster randomized trials. This approach may complicate the analysis. Although we believe that the analysis might be reasonable, more research on the statistical analysis is required to verify if the estimate of the intervention effect is not biased.

Another disadvantage not mentioned in literature is related to the timing of the informed consent procedure. When the intention was to ask informed consent from all patients at baseline, this could not be realized. The reason was that the Medical Ethics Committee of our hospital did not consider this as acceptable for patients. Therefore, patients were asked for informed consent before the switch from the control to the intervention period. The patients who entered the study after the switch were asked for informed consent before surgery. As it was impossible to ask all patients for informed consent at the outpatient ward in the few weeks before switching follow-up, letters were generated for this purpose. Consequently, patients who did not respond to the letter were not included in the intervention period, making the intervention group smaller than expected. Patients who do not respond to the letter or exit the study during the control period (e.g., because of patient death or having a recurrence) were not asked for informed consent. However, their data could still be used. This was possible as these patients did not experience any changes in follow-up and had a guaranteed anonymity (according to the Dutch law) by the assignment of unique patient numbers and a password-protected database.

2.4. DISCUSSION

Not only the stepped wedge design helps with the implementation difficulty, it is also considered more ethical because there is enough evidence
to support the efficacy of the intervention. Another advantage is that the rollout setting for all participants of the stepped wedge design motivates not only the patients but also the doctors to participate in this study. Recruitment of participants was therefore much easier in CEAwatch study. This allows hospitals to enter the control period with the same criteria of eligible patients and include new patients during the study period. This advantage has not been emphasized in literature yet. Furthermore, the sequential introduction of the intervention was a real benefit to the CEAwatch study. It would have been almost impossible to select another trial design. Other generally accepted advantages such as opportunity of time effect investigation and reduction to the number of clusters did not have the expected benefit for the study. On the other hand, the application of the stepped wedge design to CEAwatch increased the complexity in data analysis and the repeated measurements may bring additional burden to patients in terms of quality of life, satisfaction, and costs. In addition, there are concerns regarding the procedure of informed consents. This trial-specific disadvantage of steped wedge design is new to those general ones.

Because the analysis of the study is still in progress, whether using a stepped wedge design provides unbiased estimation of the treatment effect remains to be further investigated. To the extent that missing data are negligible, we believe with proper analysis method that the estimation should be adequately unbiased.

We were able to include a large number of patients. In many surgical trials, the inclusion of patients is one of the key problems. This is also a good solution to the challenge of recruitment difficulties mentioned in surgical studies. This challenge is mainly due to the strong preferences of patients and surgeons for one intervention and the organization of the inclusion and randomization.[5]

Because the original sample size calculation did not take into account
the design effect of the stepped wedge design, a more correct sample size calculation by Hussey and Hughes [10] was performed retrospectively. The minimal number of patients per cluster per time interval was determined at 187. We used the same input information that was described earlier.

The analysis of the benefits and drawbacks brought by the stepped wedge design indicates that it is a strong alternative for pragmatic cluster randomized trials such as the CEAwatch. The general advantages of the stepped wedge design still holds compared with other controlled trial design, whereas most of the general concerns regarding the stepped wedge design bring no disadvantages to the CEAwatch study. However, the stepped wedge design makes the analysis of the trial rather complex and whether repeated measurements bring burden to patients needs further investigation. One advantage that has not been mentioned in literature before is that the stepped wedge design contributes to larger sample size because of not only the ethical advantage of the design but also the rollout setting, which provides strong motivation for doctors. This allows hospitals to enter the control period with same criteria of eligible patients and include new patients during the study period. Meanwhile, difficulty in the informed consents was found as a disadvantage specifically in our clinical application.

**REFERENCES**


