INTRODUCTION
1.1. Chapter outline

The randomized controlled trial or randomized clinical trial is considered the gold standard for establishing efficacies and effectiveness of a new intervention in the medical field. In a randomized controlled trial, participants are randomly allocated to two or more treatments and data is collected from those treatment arms for comparison. A randomized controlled trial ensures that participants in different treatment arms only differ in terms of the treatment they receive and therefore difference in the outcomes can be attributed to the difference in treatments.

Randomization can be performed either at an individual level or at a cluster level. A randomized controlled trial with randomization at a cluster level is often named a clustered randomized trial. The advantage of randomization at a cluster level over randomization at an individual level is that it protects the trial from possible contamination, and that it is easier to perform when individual randomization is not feasible. However, it is more difficult to maintain balance in possible confounders for different treatment arms since participants from the same clusters will be assigned to the same treatment.

In a randomized controlled trial different clinical trial designs may be applied, including the well-known parallel group design. Other types of clinical trial design such as crossover design, factorial design are also frequently applied. Randomized controlled trials may also apply a sequential introduction of a new treatment to determine its efficacy or effectiveness with respect to a control treatment. In contrast with the classical parallel group design where different treatments are assigned to distinct groups of patients, sequential introduction of the treatment usually applies the two different treatments to the same (group of) patients or clusters in a chronological order. Such design provides opportunities for within-subject or within-cluster comparisons in addition to the between-subject
or between-cluster comparison, creating a design where patients or clusters could be considered as their own control. This type of clinical trial design is the so-called stepped wedge design.[16] This will be the topic of this thesis entitled “Evaluation and analysis of stepped wedge designs: Application to colorectal cancer follow-up” in which the epidemiological practice and application of the stepped wedge design will be discussed.

The arguments for the application of a stepped wedge design, factors to consider when designing a trial using a stepped wedge design, and the statistical analysis of data obtained from a stepped wedge design will be demonstrated on the basis of the CEA-Watch study.[32] The CEA-Watch study is a clinical trial investigating an intensification of the follow-up protocol for colorectal cancer as compared to care as usual follow-up in patients after surgical resection of their primary tumor. In the introduction to the present thesis, several aspects of the stepped wedge design and the CEA-Watch study will be introduced and discussed briefly. The chapter starts with a general introduction to the stepped wedge design, its primary merits and limitations, followed by a short historical overview of the application of the stepped wedge design in the field of cancer epidemiology and some descriptions of the CEA-Watch study. Afterwards, the application of the stepped wedge design to the motivating example, i.e., the CEA-Watch study will be outlined. This chapter will end with the aims and outline of this thesis.

1.2. STEPPED WEDGE DESIGN

A stepped wedge design is a randomized controlled trial design that utilizes sequential roll-out of the intervention. [6] At the beginning of the trial, a patient or group of patients start within the control or placebo arm. Switching to the new intervention of interest then take place at predetermined moments. During each switch moment, a part of the control
arm crossover to the intervention arm. There are no switches from the intervention arm to the control arm. In the last time period of the trial, namely the period between the last switch moment and the end of the trial, all patients or group of patients are in the intervention arm, and exposure is exclusively to the intervention. An example of a stepped wedge design with three switch moments and two hospitals per switch moment is depicted in Figure 1.1 below.

![Figure 1.1](image)

**Figure 1.1** | A schematic of a stepped wedge design with three clusters (two hospitals per cluster) and four periods (light periods indicate control and dark periods indicate intervention)

Randomization in the stepped wedge design is used to allocate (groups of) patients to different switch moments instead of different treatment arms. Current literature on stepped wedge design is predominated by the clustered stepped wedge trial and randomization is conducted at the cluster level (Figure 1.1). The taxonomy of the stepped wedge design is based on the type of cohorts involved in the trial: a stepped wedge
1.2. Stepped wedge design

design is considered to be cross-sectional if new patients are recruited and outcomes are measured step by step; on the other hand, if a static cohort is being followed throughout the course of the trial, this type of stepped wedge design is called longitudinal/cohort stepped wedge design; a combination of the cross-sectional and longitudinal stepped wedge design is named open cohort stepped wedge design.[15]

Though there are several reasons for adopting a stepped wedge design, there are three major motivations.[18, 27, 28] First of all, a stepped wedge design provides logistic conveniences and flexibilities when implementation of the new intervention cannot be realized simultaneously across all trial clusters. Such problem frequently rises from large-scale pragmatic trials [29] such as multi-center trials. To name a few causes of the difficulty, setups of the new intervention or learning the new techniques requires certain amount of time, or administration approval procedure needs different amount of time to be acquired among different locations. Under these circumstances, a stepped wedge design becomes particularly attractive since the new intervention does not need to be deployed concurrently. The second most common motivation is ethical considerations when withholding a new treatment for part of the cohort is considered unacceptable. This argument is even stronger when the efficacy of the treatment has already been demonstrated and proven. On the other hand, the ethical benefit is only true for a longitudinal or open cohort stepped wedge design, where all patients will eventually be exposed to the new intervention treatment. However, the cross-sectional stepped wedge design does not have this benefit since patients in a cross-sectional stepped wedge design adhere the same treatment as the treatment of their enrolled time period and cluster. Nonetheless, a consequence of the second motivation is the attraction of more participants and accessibility to a larger sample size. Last but not least, the stepped wedge design also provides statistical efficiency under certain conditions. For instances, it has
been shown that a stepped wedge design is more efficient in terms of estimating the treatment effect compared to the classical parallel group design when the intraclass correlation is large. An intuitive explanation is that, in a stepped wedge design, patients can be considered as their own control and therefore a stepped wedge design reduces the variations.

1.3. Sample Size Calculation

Sample size calculation for stepped wedge design shares much similarity with traditional clustered randomized controlled trial.[8, 9] In this approach, the required sample size for a parallel group design will be estimated and the estimated sample size will be multiplied by the design effect of the stepped wedge design to obtain the required sample size. For the stepped wedge design, the design effect is the ratio of the treatment effect estimator variance of the stepped wedge design and the parallel group design. This approach is model dependent, as for different models and assumptions, the design effect will differ. In current literature, the design effect is available for cross-sectional and longitudinal stepped wedge design with certain variance components model for normally-distributed outcomes.[12, 14, 20, 34] An alternative approach is to directly calculate the variance of the test statistic and obtain the required sample size assuming the test statistics is asymptotically normal.[19] Nevertheless, both sample size calculation approaches rely heavily on obtaining the variance of the estimator or test statistic. For complex models and non-normal outcomes, this may not be feasible and options are limited to simulation-based calculation [2] except for binomial outcomes for which approximation by normal distribution may be adopted.
The root of using stepped wedge design in cancer epidemiology traces back to early 1990’s. In the 1983 report of WHO meeting on the prevention of liver cancer [35], several points on designing large-scale studies to evaluate the effectiveness of immunization in preventing hepatocellular carcinoma were mentioned. It was suggested that ethical problems would be present for traditional randomized controlled trial design since vaccination of children with hepatitis B vaccine at birth would confer long-term protection against the development of hepatocellular carcinoma. Another problem was related to the costs and limited supplies of the vaccine. It was reported that for a potential trial at multiple sites, it would be unlikely to have sufficient vaccine available for all sites, especially for high-risk regions such as Africa and Asia. But considering the long follow-up time of such trial, it might be the case that the vaccine would become cheaper and more widely available during the course of the trial. Motivated by these problems, the so-called “Gambia Hepatitis Intervention Study” was conducted in The Gambia in 1987 [11] which is considered the earliest stepped wedge design trial recorded in literature. In this study, the hepatitis B virus vaccination was introduced to the “Extended Program of Immunization” in Gambia at approximately 10- to 12-week intervals by vaccination teams. All new born children recorded at the vaccination points served by the team was included. Vaccination effect was evaluated through a long-term follow-up of these children during the trial period and incidence of hepatocellular carcinoma and chronic liver diseases was compared among vaccinated children and those who were not in each 3 months period.

More recent accounts for stepped wedge design used in cancer epidemiology related trials, started to appear after the seminal paper from Hussey and Hughes in 2007 [20] and the resurface of stepped wedge design
in clinical trials in general. In a review paper on breast cancer screening [10], it was suggested that multiple time series data might be particularly useful in evaluating screening introduced across health systems or countries at different times. The stepped wedge design was proposed as one of the randomized trial design options. In the same days, trials with a focus on psychological outcome in cancer patients often used pre/post comparisons within a randomized controlled trial framework. [17, 26] This type of evaluation method (or its variation) can also rise from a stepped wedge design when measurements of psychological variables are performed at multiple time periods during the trial. Essentially, pre-/post comparison can be viewed as one form of the stepped wedge design by using one single switch moment. This has also been suggested by a systematic review on improving quality of care for lung cancer. [36] Other topics of cancer-related trials which used the stepped wedge design vary from medical education for general practitioners [31], healthcare for cancer patients [1, 4, 5], to community-based care support to primary colorectal cancer diagnosis using immunochemical faecal occult blood test. [22] The two most frequent mentioned rationales for using a stepped wedge design are ethical considerations and logistic limitations which is consistent with the general motivation as outlined above.

1.5. The CEA-Watch Trial

In this thesis, the primary motivating example is the CEA-Watch study. [32] It is a multi-center clustered stepped wedge trial conducted in 11 non-academic teaching hospitals in the Netherlands between the period of 2010 and 2012. The objective of the study was to investigate whether intensification of the follow-up protocol would be associated with a higher percentage of early stage recurrences and an increased survival time in patients with recurrences as compared to care as usual follow-up protocol.
Thereby, developing an evidence-based guideline for routine follow-up using standard screening tools, namely CEA, short for carcinoembryonic antigen, and imaging techniques such as computer tomography. Eligible participants were colorectal cancer patients with American Joint Committee on Cancer (AJCC) stage I-III after R0 resection whom had been operated from 2007 until July 2012. Patients who were not medically fit for metastasectomy, diagnosed with other malignancies or had metachronous metastases at the start of the trial were excluded. The intervention follow-up protocol adhered to CEA measurements every two months and yearly imaging in patients’ first three years of follow-up. Outpatient clinic visits with imaging of chest and abdomen were scheduled on a yearly basis in the same period. In case of more than 20% increase in CEA value with absolute value higher than 2.5 ng/ml, another blood sample was drawn. If a consecutive rise was observed, a CT scan of chest and abdomen was advised. The care as usual follow-up protocol following the 2008 national guideline of the Netherlands was considered as the control arm and is consisted of outpatient clinical visits every 3-6 months in the first three years of patients’ follow-up and every single year in the last two years. A comparison between the two follow-up protocols is shown in Figure 1.2. The primary outcome of the trial was the percentage of curative treated recurrences among all patients which are considered at risk for developing recurrences. Secondary outcomes included time-to-detection of the recurrences, quality of life and mental wellness of the patients, and long term survival for patients with recurrences.

The eleven participating hospitals were grouped into five clusters, three smaller hospitals were grouped together to ensure clusters were balanced in terms of the number of patients. Every three months, a cluster switched from the care as usual protocol to the intervention protocol until all clusters were switched. The order of the switch is randomized using simple randomization. Substantial overlapping between the recruiting period
1.6. STEPPED WEDGE DESIGN FOR CEA-WATCH

The rational for adopting a stepped wedge design in the CEA-Watch trial will be briefly discussed here. As the use of CEA for colorectal cancer follow-up have been established [7, 21, 23, 30, 33], the next step was to evaluate its effectiveness and cost-effectiveness on the community level. Randomized and controlled experimental approaches might have the highest internal validity, but are not always feasible and are difficult and
1.6. **Stepped wedge design for CEA-Watch**

Costly to implement in larger population to provide support for evidence-based decision making. Moreover, these are often not best suited for testing complex interventions with long term lifestyle changes.[25] As the CEA-Watch study promoted a systematic evaluation of a complex follow-up protocol on its pragmatic nature, stepped wedge design was appealing for the investigators when planning the trial. As a prerequisite, the CEA-Watch trial required a computer assisted system [13] to be installed and functioning at each participating hospital to ensure the validity and adherence of the follow-up under study. It was deemed unrealistic to start multiple hospitals with the new intervention at the same time. In addition, approvals from each local medical ethic committees were anticipated to be cumbersome. Thus the stepped wedge design became one of the better choices for its staggered starting points.

However, adopting a relatively new trial design such as the stepped wedge design also imposed challenges and problems. From a designing prospective, some of the limitations have already been foreseen during the planning phase. For instance, it could be expected that there would be an increased risk of attrition due to the prolonged waiting time for the new intervention as well as an increased risk of contamination. These potential problems put doubtful clouds above the trial validity and biasness [24], and therefore requires prudent examination. On the other hand, from a statistical or data analysis perspective, resources and information with regards to analyzing different endpoint outcomes under the contexts of the stepped wedge design are limited in literature. Thus far, only the method for analyzing continuous normally-distributed outcome has been addressed.[12, 20] It is unknown for other types of outcomes, such as relative risk or survival time, whether the traditional statistical methods would be still appropriate. If not, then what kind of adjustment is needed? Especially in the CEA-Watch trial, the primary outcome is a relative risk type of outcome and it also has survival time and questionnaire as its
secondary outcomes. Therefore, there is a need to investigate on these questions and demonstrate the proper methods.

1.7. Thesis aims and outline

1.7.1. Aim

The aim of the thesis is to investigate the practice of the stepped wedge design for epidemiological studies, specifically large-scale pragmatic clinical trials. Furthermore, to discuss and illustrate the appropriate data analysis methods for various types of outcomes commonly seen from such trials.

1.7.2. Outline

In the first part of the thesis, the focus is on the theoretical discussion of the stepped wedge design. In Chapter 2, its common strengths and weaknesses are surveyed and summarized. The merits of the application of a stepped wedge design specific to the CEA-Watch study is closely examined and discussed. The results demonstrate that not all perceived traits of the stepped wedge design apply to the CEA-Watch study. The implications from this chapter can be generalized to more situations similar to the CEA-Watch study. In Chapter 3, data analysis methods for the stepped wedge design are discussed. It demonstrates the usage of different methods for different outcome types and highlights the assumptions made by these methods, when to use or not use such method, and what are the caveats to consider when analyzing the data that arise from a stepped wedge design trial.

The second part of the thesis consists of three examples of data analyses from the CEA-Watch study. Chapter 4 considers the main outcome of interests, the proportion of recurrences with curative treatment as well as the time-to-detection of the recurrences. This proportion can be con-
considered as a binary outcome and the time-to-detection is a survival time type of variable. The difficulty lies in the fact that patients switch treatment in the stepped wedge design so treatment need to be considered as time-dependent. In Chapter 5, the long term survival time is evaluated for patients that have developed recurrence during the trial. It is necessary to distinguish this with the survival time analysis showed in the first example. Because once patients’ recurrences have been detected, they belong to a specific treatment group based on the detection method, and the “treatment” is no longer time-dependent as in the first case. The last example shown in Chapter 6 is concerned with patients’ quality of life and mental well-being during the trial. As a secondary outcome, the questionnaires were only filled out by patients at two distinct time points during the trial and an ANOVA-type model is used to make sensible inferences from the data.

Finally, Chapter 7 contains a summary and general discussion on the results of the thesis and discusses generalization and prospective future research.

REFERENCES


open 5(7):e008,921


[18] de Hoop E, van der Tweel I, van der Graaf R, Moons KG, van Delden JJ, Reitsma JB, Koffijberg H (2015) The need to balance merits and limitations from different disciplines when considering the stepped wedge clus-
REFERENCES


[27] Mdege ND, Man MS, Taylor CA, Torgerson DJ (2011) Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. J Clin Epidemiol 64(9):936–948

[28] Mdege ND, Man MS, Taylor CA, Torgerson DJ (2012) There are some circumstances where the stepped-wedge cluster randomized trial is preferable to the alternative: no randomized trial at all. response to the commentary by Kotz and colleagues. J Clin Epidemiol 65(12):1253


[33]


