Global Epidemiology 8 (2024) 100163

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology

Methodology

Tailored guidance to apply the Estimand framework to Trials within Cohorts (TwiCs) studies

R. Gal [a,1,*], R. Kessels [b,1], K. Luijken [c], L.A. Daamen [a], D.R. Mink van der Molen [a], S.A.M. Gernaat [a], A.M. May [c], H.M. Verkooijen [a], P.M. van de Ven [b]

[a] Division of Imaging and Oncology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[b] Julius Center for Health Sciences and Primary Care, Department of Data Science and Biostatistics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[c] Julius Center for Health Sciences and Primary Care, Department of Epidemiology and Health Economics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Trials within Cohorts study
TwiCs
Cohort multiple randomized controlled trial
cmRCT
Estimand
Intercurrent events

ABSTRACT

Objective: The estimand framework offers a structured approach to define the treatment effect to be estimated in a clinical study. Defining the estimand upfront helps formulating the research question and informs study design, data collection and statistical analysis methods. Since the Trials within Cohorts (TwiCs) design has unique characteristics, the objective of this study is to describe considerations and provide guidance for formulating estimands for TwiCs studies.

Methods: The key attributes of an estimand are the target population, treatments that are compared, the endpoint, intercurrent events and their handling, and the population-level summary measure. The estimand framework was applied retrospectively to two TwiCs studies: the SPONGE and UMBRELLA Fit trial. The aim is to demonstrate how the estimand framework can be implemented in TwiCs studies, thereby focusing on considerations relevant for defining the estimand. Three estimands were defined for both studies. For the SPONGE trial, estimators were derived.

Results: Intercurrent events considered to occur exclusively or more frequently in TwiCs studies compared to conventional randomized trials included intervention refusal after randomization, misalignment of timing of routine cohort measurements and the intervention period, and participants in the control arm initiating treatments similar to the studied intervention. Considerations for handling refusal after randomization related to decisions on whether the target population should include all eligible participants or the subpopulation that would accept (or undergo) the intervention when offered. Considerations for handling treatment initiation in the control arm and misalignments of timing related to decisions on whether such events should be considered part of treatment policy or whether interest is in a hypothetical scenario where such events do not occur.

Conclusion: The TwiCs study design has unique features that pose specific considerations when formulating an estimand. The examples in this study can provide guidance in the definition of estimands in future TwiCs studies.

What is new?

What are the key findings?

- The TwiCs study design has unique features that pose specific considerations when formulating an estimand. One of these features is that patients randomized to the intervention arm are asked second-stage consent to accept the intervention.

What does this add to what is known related to methods research within the field of clinical epidemiology?

- Three example estimands were retrospectively defined for each of two previously conducted TwiCs studies, the SPONGE trial and the UMBRELLA Fit trial, which provide guidance for formulating estimands for future TwiCs studies.

* Corresponding author at: Division of Imaging and Oncology, University Medical Center Utrecht, PO Box 85500, 3508 GA, Utrecht, the Netherlands.
E-mail address: R.Gal@umcutrecht.nl (R. Gal).
[1] Shared first authorship

https://doi.org/10.1016/j.gloepi.2024.100163
Received 21 June 2024; Received in revised form 16 September 2024; Accepted 18 September 2024
Available online 21 September 2024

- For the SPONGE trial, estimators were derived for the treatment effects targeted by three estimands, all focusing on the treatment effects in different subpopulations (i.e., intention-to-treat population, both intervention and control compliers, and intervention accepters) that are potentially of interest in TwiCs studies.

What is the implication?

- Protocols of TwiCs studies should specify the targeted estimands upfront as they inform choices regarding study design, data collection and statistical analysis methods.
- Manuscripts reporting findings of TwiCs studies should report which estimands were targeted, as described in the protocol, to ensure correct interpretation of study findings.
- Not providing second-stage consent after randomization to the intervention arm (intervention refusal) should always be anticipated and considered an intercurrent event in TwiCs studies.
- In this manuscript, the discussion of estimands and corresponding estimation strategies was limited to six examples in two TwiCs studies. Future methodological research is needed to develop appropriate estimators for various estimands under varying assumptions and for different types of outcome measures.

Introduction

The Trials within Cohorts (TwiCs) design is a pragmatic design for randomized controlled trials (RCTs), generating robust and high-quality real-world evidence [1–3]. Unique features of the TwiCs design include: i) the use of an ongoing observational cohort study for participant recruitment into trials (i.e., TwiCs studies) and endpoint collection, and ii) a two-staged, patient-centered, informed consent procedure. Participants provide first-stage consent for being prospectively followed in a cohort study and potential randomization into future TwiCs studies [4,5]. Cohort participants eligible for a TwiCs study are randomized to the intervention or control arm. In contrast to standard RCTs, only participants randomized to the intervention arm are informed about the trial and asked second-stage consent to undergo the intervention, which they can accept or refuse. The TwiCs design may overcome some challenges faced in standard RCTs including slow accrual rates and early drop-out in the control arm. In addition, the TwiCs design aims to reduce contamination in the control arm by only informing those randomized to the intervention arm about the trial. Furthermore, the second-stage consent provides valuable information on acceptance of treatment [4,5].

An essential aspect of randomized intervention studies is that the research objectives and statistical analyses should be unambiguously defined upfront. The ICH E9(R1) (*Statistical Principles for Clinical Trials*) addendum on estimands provides guidance on how a research question can be translated into a treatment effect that the study aims to quantify [6]. The ICH E9(R1) addendum focuses on the definition and handling of so-called intercurrent events. In randomized studies, intercurrent events are all events that occur after randomization that may alter the course of the treatment or affect the interpretation or existence of the endpoint. Clearly defined estimands and subsequent alignment of study design, outcome collection and statistical analysis methods, ensure that study results can be used to inform decisions by relevant stakeholders, such as regulators, policy makers, care professionals and patients [7–9].

The unique features of the TwiCs design pose specific considerations when formulating an estimand. In this article, we provide example estimands for two previously performed TwiCs studies and identify the considerations for the estimand, including intercurrent events and strategies for handling. The example estimands may provide guidance when formulating estimands in TwiCs studies.

Methods

A team of statisticians and clinical epidemiologists experienced in conducting and reporting of TwiCs studies discussed challenges for estimand definition using the ICH E9(R1) addendum in the setting of TwiCs studies [6]. Intercurrent events were identified based on experiences and lessons learned from previous TwiCs studies in oncology [4,5,10]. For two selected TwiCs studies (see below), authors retrospectively defined different estimands considering the clinical context and methods to address the intercurrent events. For one TwiCs study, estimators were derived for the treatment effects targeted by three estimands (Table 1).

Key attributes of an estimand

The five key attributes of an estimand are: [7,8]

1. Target population: patients targeted by the clinical research question.
2. Treatments: the intervention(s) of interest and the comparator intervention(s), including a detailed specification of the interventions.
3. Endpoint: the outcome measure that is assessed, including at which time point or over which interval.
4. Intercurrent events: all events that may occur after randomization that alter the predefined course of the intervention(s) and/or affect the interpretation or existence of the endpoint. Researchers need to prespecify how each intercurrent event will be handled, for which the addendum defines five different strategies: treatment policy, hypothetical, composite, while on treatment, and principal stratum strategy (Table 2).
5. Population-level summary: the analytical measure used to summarize the targeted treatment effect.

TwiCs example studies

We illustrate the application of the estimand framework to TwiCs studies by post-hoc definition of estimands for the SPONGE and UMBRELLA Fit trial. These trials were selected to cover different types of interventions (i.e., a one-time point intervention and a longitudinal intervention program) and different intercurrent events, including intercurrent events occurring exclusively or more frequently in TwiCs studies compared to standard RCTs.

The SPONGE trial, nested within the Prospective Dutch Colorectal Cancer (PLCRC) cohort, investigated the effect of sponge-assisted laparoscopic surgery (i.e., use of a intraoperative retractor sponge to create a clear field of view during surgery) versus Trendelenburg laparoscopic surgery (usual care, i.e., positioned at an angle of 15 to 40 degrees with the head down) on length of hospitalization in patients undergoing surgery for sigmoid or rectal cancer [11,12]. Participants of the PLCRC cohort that were eligible for the SPONGE trial were randomized to either sponge-assisted (experimental intervention) or Trendelenburg surgery (control arm; Fig. 1). Patients randomized to the experimental intervention could accept or refuse sponge-assisted surgery, receiving

Table 1
Definition of an estimand, estimator, and estimate according to the Glossary of the ICH E9(R1) addendum on estimands [6].

Term	Definition
Estimand	"A precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarizes at a population-level what the outcomes would be in the same patients under different treatment conditions being compared."
Estimator	"A method of analysis to compute an estimate of the estimand using clinical trial data."
Estimate	"A numerical value computed by an estimator."

Table 2
Different strategies for handling intercurrent events, adapted from Ratitch et al. [8]

Strategy	Definition
Treatment policy	The intercurrent event is considered irrelevant in defining the treatment effect. The intercurrent event is taken to be part of the treatment effect of interest.
Principal stratification	The interest is in the treatment effect in a subpopulation (principal stratum) in which the intercurrent event would occur or a subpopulation in which the intercurrent event would not occur.
Hypothetical	The interest is in the treatment effect under a hypothetical scenario where the intercurrent event did not occur.
Composite variable	The intercurrent event itself is considered informative about the outcome of a participant and is incorporated into the definition of a composite outcome variable. The interest is in the effect of the treatment on this composite outcome.
While on treatment	Outcomes up to the time of the occurrence of the intercurrent event provide all necessary information about the treatment effect. The interest is in the treatment effect until occurrence of the intercurrent event regardless of actual length of treatment.

Trendelenburg surgery in case of refusal. However, if the physician decided during surgery that only using Trendelenburg position or the retractor sponge was not sufficient, the physician used both the retractor sponge and the Trendelenburg position during the same surgical procedure [13].

The UMBRELLA Fit trial, nested within the Utrecht cohort for Multiple Breast cancer intErvention studies and Long-term evaLuAtion (UMBRELLA) [14], evaluated the effect of a 12-week exercise program compared to usual care on quality of life (QoL) at six months after randomization in women who completed breast cancer treatment [15,16]. Women were randomized at 12 or 18 months after they started radiotherapy treatment, and the cohort measurement at 18 or 24 months cohort follow-up (i.e., six months later) was used as endpoint for effect estimation.

Results

SPONGE trial

Three intercurrent events were defined:

- Intercurrent event 1: No second-stage consent, i.e., refusal of sponge-assisted surgery after randomization to sponge-assisted surgery (intervention arm)
- Intercurrent event 2: Use of both the Trendelenburg position and retractor sponge at the discretion of the treating physician (both arms)
- Intercurrent event 3: Any event unrelated to the studied interventions that leads to prolonged hospital stay, e.g. a viral infection acquired post-surgery (both arms)

Three estimands varying in target population, treatment definition and handling of the intercurrent events were defined and the estimator for the treatment effect targeted by the estimand was described (Table 3). In all three estimands, intercurrent event 3 is handled under a treatment policy strategy, meaning that comparison of interest concerns the total duration of hospital stay, irrespective of any prolongations of hospital stay that are unrelated to the studied interventions. For handling intercurrent events 1 and 2, either the treatment policy strategy (estimand 1), principal stratum strategy (estimand 2) or a combination of the two strategies (estimand 3) is proposed. Under the principal stratum strategy, the treatment effect is estimated within a specific subpopulation (stratum). Estimand 1, 2 and 3 all concern effects in different (sub)populations as shown in Fig. 2.

SPONGE estimand 1

One objective of the SPONGE trial could be to evaluate whether offering sponge-assisted surgery reduces the length of postoperative hospital stay compared to initiating Trendelenburg surgery in the eligible study population (Table 3). This generates evidence about the efficacy of offering sponge-assisted surgery, regardless of whether patients accept and undergo this intervention, and is informative for clinical practice when both Trendelenburg and sponge-assisted surgery are treatment options. Estimand 1 therefore considers the complete study population

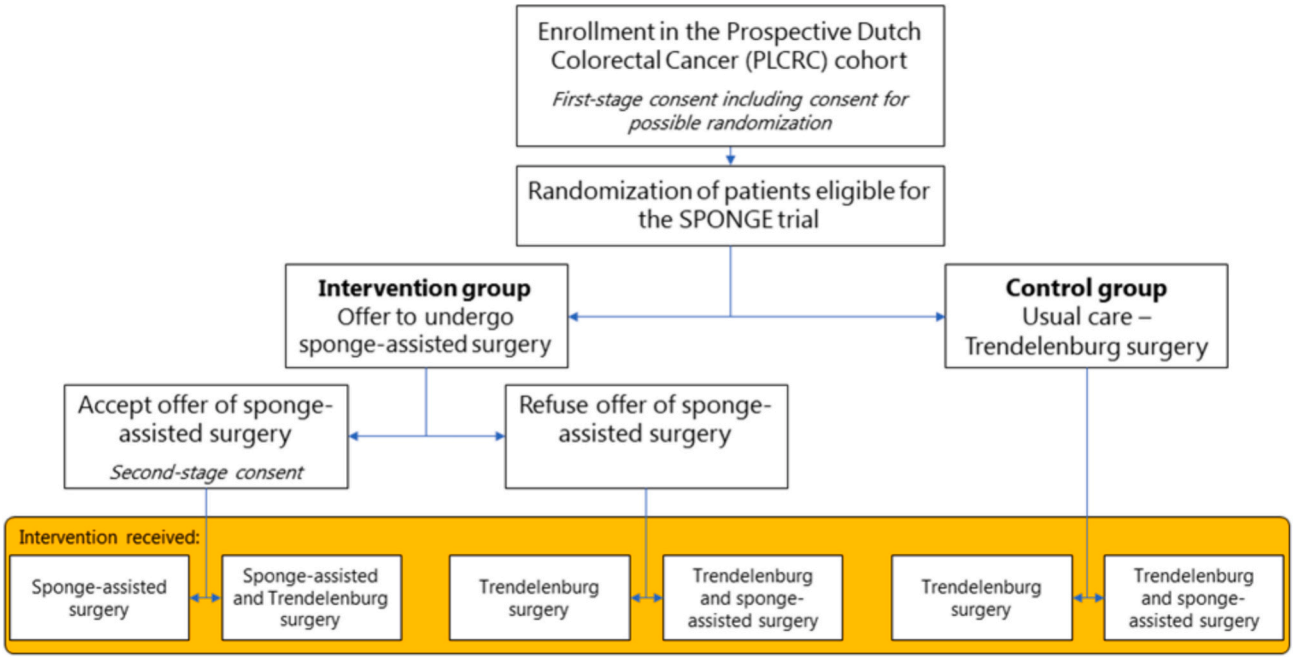

Fig. 1. Flowchart of the SPONGE trial.

Table 3
Overview of estimands that were defined retrospectively for the SPONGE trial.

Estimand 1

Population	Patients with sigmoid or rectal cancer planned for elective colorectal laparoscopic surgery with WHO 0–2 (intention-to-treat population)
Treatments	*Intervention*: Sponge-assisted surgery, but additional Trendelenburg surgery at physician's discretion is allowed (IE2) *Control*: Trendelenburg surgery, but additional sponge-assisted surgery at physician's discretion is allowed (IE2)
Strategy to handle intercurrent events	IE: 1–3; Strategy: Treatment policy; Description: Ignoring IE1–3 when estimating the effect
Research question	What is the effect of offering sponge-assisted surgery compared to Trendelenburg surgery on the length of hospital stay in patients with sigmoid or rectal cancer planned for elective colorectal laparoscopic surgery with WHO 0–2?

Estimand 2

Population	Patients with sigmoid or rectal cancer planned for elective colorectal laparoscopic surgery with WHO 0–2, who would accept and only undergo sponge-assisted surgery when offered and who would only undergo Trendelenburg surgery when this is usual care (both intervention and control compliers)
Treatments	*Intervention*: Sponge-assisted surgery, not allowing additional Trendelenburg surgery at physician's discretion (IE2) *Control*: Trendelenburg surgery, not allowing sponge-assisted surgery at physician's discretion (IE2)
Strategy to handle intercurrent events	IE: 1–2; Strategy: Principal stratum; Description: Restricting to patients where IE1–2 would not occur (i.e., both intervention and control compliers) IE: 3; Strategy: Treatment policy; Description: Ignoring IE3 when estimating the effect
Research question	What is the effect of sponge-assisted surgery compared to Trendelenburg surgery on the length of hospital stay in patients with sigmoid or rectal cancer planned for elective colorectal laparoscopic surgery with WHO 0–2, in the subpopulation of patients that would accept and only undergo sponge-assisted surgery when offered and who would undergo only Trendelenburg surgery when this is usual care?

Estimand 3

Population	Patients with sigmoid or rectal cancer planned for elective colorectal laparoscopic surgery with WHO 0–2, who would accept sponge-assisted surgery when offered (intervention accepters)
Treatments	Similar to estimand 1
Strategy to handle intercurrent events	IE: 1; Strategy: Principal stratum; Description: Restricting to patients where IE1 would not occur (i.e., intervention accepters) IE: 2–3; Strategy: Treatment policy; Description: Ignoring IE2–3 when estimating the effect
Research question	What is the effect of sponge-assisted surgery compared to Trendelenburg surgery on the length of hospital stay in patients with sigmoid or rectal cancer planned for elective colorectal laparoscopic surgery with WHO 0–2, in the subpopulation of patients who would accept (but not necessarily only undergo) sponge-assisted surgery when offered?

Abbreviations: IE, intercurrent events; WHO, World Health Organization.
Common elements for all three estimands not explicated in the table columns are as follows:
Intercurrent events:
- IE1: No second-stage consent, i.e., refusal of sponge-assisted surgery after randomization to sponge-assisted surgery (intervention arm)
- IE2: Use of both the Trendelenburg position and retractor sponge at the discretion of the treating physician (both arms)
- IE3: Any event unrelated to the studied interventions that leads to prolonged hospital stay, e.g. a viral infection acquired post-surgery (both arms)

Endpoint: Length of postoperative hospital stay, defined as days from surgery until discharge.
Population-level summary: Mean difference.

and treatments as initially offered, irrespective of deviations from the offered treatment. This estimand targets the intention-to-treat (ITT) effect where intercurrent events 1 and 2 are handled using the treatment policy strategy and therefore outcomes are used regardless of whether intercurrent events occurred or not.

SPONGE estimand 2

Patients may refuse sponge-assisted surgery when offered and physicians may decide to use both Trendelenburg and sponge-assisted surgery. Both these intercurrent events can mitigate the treatment effect. An alternative objective may be to understand the effect on the length of hospital stay in the (sub)population of patients that would only undergo the intervention to which they were randomized (Table 3). Estimand 2 considers the effect of only sponge-assisted surgery compared to only Trendelenburg surgery on the length of hospital stay in the subpopulation who, when offered, would accept and undergo sponge-assisted surgery only, but who would undergo Trendelenburg surgery only when Trendelenburg surgery is the usual care (both intervention and control compliers). The population of interest is defined as the subpopulation of patients where both intercurrent events 1 and 2 do not occur. This corresponds to handling these intercurrent events using the principal stratum strategy.

SPONGE estimand 3

Another objective may be to estimate the effect of sponge-assisted surgery on length of hospital stay in those that initially accept sponge-assisted surgery when offered, allowing it to be combined with Trendelenburg if needed. The corresponding effect estimate is particularly valuable from a patient's and health-care provider's perspective as the decision to accept the intervention occurs before undergoing the intervention. Estimand 3 considers the effect of sponge-assisted surgery compared to Trendelenburg surgery on the length of hospital stay in the subpopulation of patients who would accept sponge-assisted surgery when offered, irrespective of whether they undergo only sponge-assisted surgery or in combination with Trendelenburg surgery (and irrespective of what they would undergo under Trendelenburg; Table 3). In this case, intercurrent event 1 is handled using the principal stratum strategy restricting to the patients that would accept sponge-assisted surgery when offered, whereas intercurrent event 2 is handled using the treatment policy strategy (accepting adding sponge-assisted surgery to Trendelenburg surgery and vice versa).

Estimation of the treatment effect for SPONGE estimands 1–3

Fig. 2 illustrates how the treatment effects for the three estimands can be estimated. Estimand 1 handles all intercurrent events using a treatment policy strategy. Estimation therefore follows the standard ITT principles and should include all randomized patients in the treatment

A. Patient strata within the SPONGE trial

			Randomized to Trendelenburg surgery: Undergo Trendelenburg and sponge-assisted surgery*	Randomized to Trendelenburg surgery: Undergo Trendelenburg surgery (*control compliers*)	Marginal proportions
Randomized to sponge-assisted surgery	No second-stage consent (*intervention refusers*)	Undergo Trendelenburg surgery	(1) Non-existent	(2) Proportion: $1-p_1-p_2-p_3$ and effect: 0	$1-p_1-p_2-p_3$
		Undergo Trendelenburg and sponge-assisted surgery*	(3) Proportion: p_1 and effect: 0	(4) Non-existent	p_1
	Second-stage consent given (*intervention accepters*)	Undergo sponge-assisted and Trendelenburg surgery*	(5) Proportion: $p_2 = q_1 - p_1$ and effect: 0	(6) Non-existent	p_2
		Undergo sponge-assisted surgery (*intervention compliers*)	(7) Non-existent	(8) Proportion: p_3 and effect: CACE	p_3
		Marginal proportions	q_1	$1-q_1$	1

B. Estimand 1: Randomized population

Estimator: $\hat{\mu}_{ITT} = \bar{y}_{\text{randomized to sponge}} - \bar{y}_{\text{randomized to Trendelenburg}}$

			Randomized to Trendelenburg: Undergo Trendelenburg and sponge-assisted surgery*	Randomized to Trendelenburg: Undergo Trendelenburg surgery (*control compliers*)
Randomized to sponge-assisted surgery	No second-stage consent (*intervention refusers*)	Undergo Trendelenburg surgery		
		Undergo Trendelenburg and sponge-assisted surgery*		
	Second-stage consent given (*intervention accepters*)	Undergo sponge-assisted and Trendelenburg surgery*		
		Undergo sponge-assisted surgery (*intervention compliers*)		

C. Estimand 2: Intervention and control compliers

Estimator: CACE = $\hat{\mu}_{ITT} / \hat{p}_3$

			Randomized to Trendelenburg: Undergo Trendelenburg and sponge-assisted surgery*	Randomized to Trendelenburg: Undergo Trendelenburg surgery (*control compliers*)
Randomized to sponge-assisted surgery	No second-stage consent (*intervention refusers*)	Undergo Trendelenburg surgery		
		Undergo Trendelenburg and sponge-assisted surgery*		
	Second-stage consent given (*intervention accepters*)	Undergo sponge-assisted and Trendelenburg surgery*		
		Undergo sponge-assisted surgery (*intervention compliers*)		

D. Estimand 3: Intervention accepters

Estimator: $\hat{\mu}_{ITT} / (\hat{p}_2 + \hat{p}_3)$

			Randomized to Trendelenburg: Undergo Trendelenburg and sponge-assisted surgery*	Randomized to Trendelenburg: Undergo Trendelenburg surgery (*control compliers*)
Randomized to sponge-assisted surgery	No second-stage consent (*intervention refusers*)	Undergo Trendelenburg surgery		
		Undergo Trendelenburg and sponge-assisted surgery*		
	Second-stage consent given (*intervention accepters*)	Undergo sponge-assisted and Trendelenburg surgery*		
		Undergo sponge-assisted surgery (*intervention compliers*)		

Fig. 2. Patient strata for the estimands of the SPONGE trial and how the treatment effects can be estimated. The subpopulations or strata that are defined as target population in the estimand are represented by the yellow cells. **Panel A**: Each cell represents a (sub)population/stratum which is denoted by (1)–(8). For each of the eight strata, the corresponding proportion as part of the randomized population and causal treatment effect are described. The following assumptions are made: i) the treatment effect is zero in cells (2), (3) and (5) where patients undergo the same treatment irrespective of randomization and ii) patients that need to undergo both treatments do so irrespective of randomization and second-stage consent, implying that cells (1), (4), (6) and (7) are empty. **Panel B**: Estimand under the treatment policy strategy (targeting ITT effect). **Panel C**: Estimand for the principal stratum of patients who would always comply to their assigned treatment. **Panel D**: Estimand for the principal stratum of patients who would accept sponge-assisted surgery when offered. *Abbreviations*: ITT, intention-to-treat; CACE, Complier Average Causal Effect. *At physician's discretion. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

arm to which they were randomized and all the outcomes as actually observed, irrespective of any prolongations of hospital stay unrelated to the interventions studied (Fig. 2B).

Principal stratum strategies as considered in estimand 2 and 3 restrict the population based on the intercurrent events that occur post-randomization. Fig. 2A decomposes the population into eight strata as defined by intercurrent event 1 and 2. The treatment effects under the principal stratum strategy defined for the SPONGE trial can be estimated when making the following assumptions: [17,18] i) the treatment effects in strata (2), (3), and (5), where patients undergo the same (combination of) treatments irrespective of the randomization, are always zero (their outcome would be equal under both treatment arms); ii) patients that require both Trendelenburg and sponge-assisted surgery always do so, irrespective of the treatment that is first initiated, implying that strata (1), (4), (6) and (7) are empty; and iii) there are no missing outcome data. Fig. 2C shows the principal stratum for estimand 2. Under the stated assumptions, the treatment effect targeted by estimand 2 can be estimated by dividing the ITT estimator by the complier fraction defined

as the proportion of patients randomized to sponge-assisted surgery who accept and only undergo sponge-assisted surgery. The estimator is a Complier Average Causal Effect (CACE) estimator as it targets the treatment effect for the subpopulation of patients that would be fully compliant in both treatment arms. Fig. 2D shows the principal stratum for estimand 3. Under the stated assumptions, the treatment effect targeted by estimand 3 can be estimated by dividing the ITT estimator by the proportion of patients randomized to sponge-assisted surgery who accept the intervention. It should be noted that treatment effect estimates under the principal stratum strategy are well-defined causal effects, since principal stratum membership is not affected by treatment assignment [19]. For further reading on how to use principal stratification in analysis of clinical trials, see Lipkovich et al. (2022) [18].

UMBRELLA Fit trial

Six intercurrent events were identified:

- Intercurrent event 1: No second-stage consent, i.e., refusal of the exercise program after randomization to the intervention arm (intervention arm)
- Intercurrent event 2: Delay in start of the 12-week exercise program due to which the program is still ongoing six months after randomization (intervention arm)
- Intercurrent event 3: The 12-week exercise program not (yet) started within six months after randomization despite second-stage consent (intervention arm)
- Intercurrent event 4: Early withdrawal from the exercise program (intervention arm)
- Intercurrent event 5: Participating in an exercise program at own initiative (control arm)
- Intercurrent event 6: Increasing physical activity and exercise at own initiative other than by participating in an exercise program (control arm)

Table 4
Overview of estimands that were defined retrospectively for the UMBRELLA Fit trial.

Estimand 1	
Population	Women 18–75 years old who completed primary breast cancer treatment, have a physically inactive lifestyle, and are 12–18 months after their intake with a radiation oncologist (intention-to-treat population)
Treatments	*Intervention*: 12-week exercise program, allowing delayed start (IE2 and IE3) and early withdrawal (IE4) *Control*: Usual care, allowing participation in exercise programs (IE5) and increased physical activity at own initiative (IE6)
Strategy to handle intercurrent events	IE: 1–6 — Strategy: Treatment policy — Description: Ignoring IE1–6 when estimating the effect
Research question	What is the effect of offering the 12-week exercise program compared to usual care on QoL after 6 months in women 18–75 years old who completed primary breast cancer treatment, have a physically inactive lifestyle, and are 12–18 months after their intake with a radiation oncologist?
Estimand 2	
Population	Similar to estimand 1 (intention-to-treat population)
Treatments	*Intervention*: 12-week exercise program, not allowing delayed start (IE2 and IE3) and early withdrawal (IE4) *Control*: Usual care, not allowing participation in exercise programs (IE5) and allowing increased physical activity at own initiative (IE6)
Strategy to handle intercurrent events	IE 1–4: Strategy Hypothetical — Description: As if IE1–4 did not occur (i.e., as if all women in the intervention arm completed the exercise program within 6 months) IE 5: Strategy Hypothetical — Description: As if IE5 did not occur (i.e., as if no women in the control arm participated in an exercise program) IE 6: Strategy Treatment policy — Description: Ignoring IE6 when estimating the effect
Research question	What is the expected effect of everyone completing the full 12-week exercise program compared to usual care without participating in exercise programs on QoL after 6 months in the population of women 18–75 years old who completed primary breast cancer treatment, have a physically inactive lifestyle, and are 12–18 months after their intake with a radiation oncologist?
Estimand 3	
Population	Women 18–75 years old who completed primary breast cancer treatment, have a physically inactive lifestyle, and are 12–18 months after their intake with a radiation oncologist, who accept the 12-week exercise program when offered (intervention accepters)
Treatments	Similar to estimand 1
Strategy to handle intercurrent events	IE 1: Strategy Principal stratum — Description: Restricting to women where IE1 did not occur (i.e., intervention accepters) IE 2–4: Strategy Hypothetical — Description: As if IE2–4 did not occur (i.e., as if accepters completed the exercise program within 6 months) IE 5: Strategy Hypothetical — Description: As if IE5 did not occur (i.e., as if no women in the control group participated in an exercise program) IE 6: Strategy Treatment policy — Description: Ignoring IE6 when estimating the effect
Research question	What is the expected effect of everyone completing the full 12-week exercise program compared to usual care without participating in exercise programs on QoL after 6 months in women 18–75 years old who completed primary breast cancer treatment, have a physically inactive lifestyle, and are 12–18 months after their intake with a radiation oncologist, in the subpopulation of women who would accept the 12-week exercise program when offered?

Abbreviations: IE, intercurrent events; QoL, quality of life.
Common elements for all three estimands not explicated in the table columns are as follows:
Intercurrent events:

- IE1: No second-stage consent, i.e., refusal of the exercise program after randomization to intervention arm (intervention arm)
- IE2: Delay in start of the 12-week exercise program due to which the program is still ongoing six months after randomization (intervention arm)
- IE3: The 12-week exercise program not (yet) started within six months after randomization despite second-stage consent (intervention arm)
- IE4: Early withdrawal from the exercise program (intervention arm)
- IE5: Participating in an exercise program at own initiative (control arm)
- IE6: Increasing physical activity and exercise at own initiative other than by participating in an exercise program (control arm)

Endpoint: QoL measured with the EORTC QLQ-C30 six months after randomization.
Population-level summary: Mean difference.

Three estimands were defined varying in target population, treatment definition and handling of intercurrent events (Table 4).

UMBRELLA Fit estimand 1

The motivation for and interpretation of UMBRELLA Fit estimand 1 is similar to SPONGE estimand 1: whether offering a 12-week exercise program improves QoL at six months after randomization compared to usual care, regardless of compliance with the assigned treatment. This estimand targets the ITT effect where all intercurrent events are handled using the treatment policy strategy and therefore outcomes are used regardless of whether intercurrent events occurred or not (Table 4). As the treatment policy strategy targets the effect in randomized women irrespective of the occurrence of an intercurrent event, continued data collection is required after refusal of the intervention or early withdrawal from the intervention. An advantage of the TwiCs design is that all randomized women still participate in the cohort with regular follow-up moments.

UMBRELLA Fit estimand 2

Developers of an exercise program might be interested in its potential maximum gain in QoL. For estimand 2 targeting this effect, a scenario is envisioned in which adherence to the 12-week exercise program is perfect in the intervention arm and not any exercise programs are followed in the control arm. Even though this might be unrealistic in clinical practice, it provides information on the maximum intervention effect in the whole study population under ideal circumstances.

In estimand 2, intercurrent events 1–5 are handled using the hypothetical strategy, assuming that all women in the intervention arm completed the 12-week exercise program within six months (i.e., before endpoint measurement), and as if no women in the control arm participated in an exercise program (Table 4).

UMBRELLA Fit estimand 3

One may be interested in the treatment effect in a subpopulation that would accept the exercise program when offered. However, also for this subpopulation, interest could still be in the maximum gain in QoL in an ideal world of complete adherence as explained in estimand 2 (as if intercurrent events 2–5 would not occur). Therefore, this estimand is similar to UMBRELLA Fit estimand 2, except that intercurrent event 1 is handled using a principal stratum strategy, restricting to a subpopulation of women that would accept the offer of the exercise program (Table 4).

Estimation strategies of the treatment effect for UMBRELLA Fit estimands 2 and 3

In estimand 2 and 3, the hypothetical strategy is proposed to handle (most) intercurrent events. The implication of the hypothetical strategy is that it involves outcomes different from those actually observed. One strategy is to consider those outcomes missing and use multiple imputation to re-estimate those outcomes to target the treatment effect under the hypothetical scenario. For example, a possible estimation strategy for estimand 3 is to first consider all post-baseline measurements of participants with intercurrent events 2–4 (intervention arm) as missing and use multiple imputation in the intervention accepters to impute outcomes under a missing-at-random assumption (as if all completed the intervention six months after randomization). Second, to handle intercurrent event 5 (control arm) with a hypothetical strategy, a similar approach can be followed using multiple imputation applied in the control arm (as if none participated in a physical exercise program). The imputation models should include both predictors for the outcome as well as predictors for the intercurrent event. Finally, an estimate for the effect in the subpopulation that would accept the intervention when offered (handling intercurrent event 1 with a principal stratum strategy) can be obtained dividing the treatment effect estimated in the randomized population (after the imputation for intercurrent events 2–5) by the proportion of women in the intervention arm who accept the intervention. The assumption underlying this final step is that the treatment effect is zero in the stratum of women who would not accept the intervention when offered, as those women would always receive the control treatment irrespective of the randomization. Alternative strategies for this final step have been described by Gal et al. and included instrumental variables (IV) and propensity score matching techniques [16]. In the propensity score matching method, the treatment effect was estimated in intervention accepters by comparing them to control patients who would have accepted the intervention if offered, whereas in the IV analysis, the relation between treatment assignment and acceptance was estimated and the predicted values were used as independent variable in a regression model estimating the effect of the exercise program on the outcome.

Estimand 2 further requires handling of outcomes for the non-accepters in the intervention arm under a hypothetical scenario as if they completed the intervention. A strategy then could be to assume all post-baseline measurements of subjects with intercurrent events 1–4 missing (intervention arm) and use multiple imputation in the intervention arm to impute outcomes under a missing at random assumption. Handling intercurrent event 5 for estimand 2 can be done as described for estimand 3. Whether this method will provide a realistic estimate of the treatment effect targeted by estimand 3 depends on the extent to which non-accepters and accepters differ and on the proportion of women that accept the intervention, as well as on the correctness of the imputation models. Therefore, sensitivity analysis based on different assumptions and imputation models should be performed to assess robustness of the estimate.

Discussion

We illustrated how the ICH E9(R1) addendum can be used to formulate estimands for TwiCs studies. The example estimands clarified that TwiCs studies have unique features leading to intercurrent events that might affect the interpretation or existence of the endpoint. The intercurrent events of refusal of the intervention after randomization to the intervention arm, and misalignment of timing of routine cohort measurements and the intervention period are considered to occur more frequently in TwiCs studies compared to conventional RCTs, which is reflected in the difference in the study design. On the other hand, the intercurrent event of participants initiating treatments similar to the studied intervention (e.g. an exercise program in the UMBRELLA Fit trial) at their own initiative as well as intercurrent events similar to intercurrent events 2 and 3 of the SPONGE trial (i.e., using both surgical techniques in one procedure, and events unrelated to study interventions that result in prolonged hospital stay) might occur in conventional RCTs as well, but these more common intercurrent events also need to be considered when defining estimands for TwiCs studies.

In TwiCs studies, not providing second-stage consent after randomization to the intervention arm (refusal of the intervention) should always be anticipated and considered an intercurrent event. When defining the target population for an estimand in a TwiCs study, a main consideration should be whether the target population is all eligible patients or a subpopulation that would accept and/or undergo an intervention when offered. The two-staged informed consent procedure provides valuable information on accepters and refusers in the intervention arm and reasons for refusal should be considered when interpreting a treatment effect under a treatment policy strategy. Furthermore, information on accepters and refusers can be used to obtain different estimators for different targeted subpopulations (principal strata), including the CACE estimator in those fully compliant under both treatment arms. The (CACE) estimators presented here, where the ITT effect is divided by the proportion intervention accepters, are only identifiable under strict assumptions. Whether these assumptions are realistic in practice is debatable. For example, assuming equal outcomes for patients undergoing both Trendelenburg and sponge-assisted surgery under both arms implies that the sequence of

surgeries does not affect the outcome. However, it can be argued that patients starting with Trendelenburg position surgery which is than combined with using the retractor sponge may have a different outcome compared to patients that first start with sponge-assisted surgery which is then combined with Trendelenburg position surgery. Other possible strategies to estimate a targeted treatment effect under a principal stratum strategy, such as the CACE, include propensity score matching and IV analyses as applied in Gal et al. or the joint exclusion restriction model or propensity score weighting [16,20]. To estimate targeted treatment effects under a hypothetical strategy, multiple imputation, propensity score weighting and IV techniques have been proposed [17,21]. For many of these methods, it is essential to capture sufficient baseline data and key demographics upon cohort enrollment as well as regularly during cohort follow-up. However, when the intervention is time-consuming or burdensome, the number of participants with intercurrent events may be high and the subsample of completers may strongly differ from non-completers. In such cases, it is unlikely that sufficient relevant data will be collected to, for instance, properly impute the data to handle the intercurrent events under the hypothetical strategy. In general, it is recommended to perform sensitivity analyses under varying assumptions and using different statistical methods to show robustness of conclusions. Further methodological research is needed to develop appropriate estimators for estimands handling intercurrent events using different strategies and varying assumptions as well as for summary measures other than differences in means and proportions such as differences in medians.

Considerations for estimands in TwiCs studies bear implications for the design and data collection. It is necessary to invest time and effort in defining estimands with all relevant stakeholders at the protocol phase of a TwiCs study [8]. In a TwiCs study, the timing of randomization and start of the intervention should be aligned with the individuals' cohort measurements. Regarding data collection, the time anchors for relevant estimands for the population should be aligned with the frequency and timing of cohort measurements. In other words, it is crucial that the follow-up schedule of a TwiCs study follows the cohort follow-up schedule, since the latter follow-up schedule can usually not be changed [5]. When starting a cohort study, it is encouraged to include sufficient measurement moments to make sure many research questions can be answered. Finally, estimation of the targeted treatment effect under a hypothetical strategy requires advanced statistical techniques. Multiple imputation, inverse probability weighting (IPW) and instrumental variable (IV) techniques have all been proposed for handling hypothetical strategies in different settings [17,21]. For many of these methods, it is essential to capture sufficient baseline data and key demographics upon cohort enrollment as well as regularly during cohort follow-up. However, when the intervention is time-consuming or burdensome, the number of participants with intercurrent events may be high and the subsample of completers may strongly differ from non-completers. In such cases, it is unlikely that sufficient relevant data will be collected to, for instance, properly impute the data to handle the intercurrent events under the hypothetical strategy. In general, for multiple imputation methods as well as for IPW and IV techniques, sensitivity analyses under varying assumptions and choices should be performed to show robustness of conclusions.

Conclusion

This article facilitated the definition of different estimands for TwiCs studies. Estimands provide a common language for clinicians, researchers, and methodologists to ensure alignment of the trial objective with the design, conduct, analysis, and interpretation of results. A TwiCs study offers an alternative approach for pragmatic trials in which the effectiveness of an intervention is investigated in a real-world setting and has unique features that pose specific considerations when formulating an estimand. Herein, it is necessary to address a clearly defined research question and relevant intercurrent events to precisely define the treatment effect to be estimated.

Funding

None.

CRediT authorship contribution statement

R. Gal: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **R. Kessels:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **K. Luijken:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **L.A. Daamen:** Writing – review & editing, Writing – original draft, Conceptualization. **D.R. Mink van der Molen:** Writing – review & editing, Formal analysis. **S.A.M. Gernaat:** Writing – review & editing, Formal analysis. **A.M. May:** Writing – review & editing, Conceptualization. **H.M. Verkooijen:** Writing – review & editing, Conceptualization. **P.M. van de Ven:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Special thanks to J.B. Reitsma for putting forward the idea of writing an article about estimands for TwiCs studies. Also thanks to G.C.M. van Baal, R.P.A. van Eijk, E. Schuit, R.K. Stellato, M.G.P. Zuidgeest, and N.P.A. Zuithoff for their insightful contributions during two brainstorm sessions in the initial phase of this project.

References

[1] Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. BMJ 2010;340:c1066. https://doi.org/10.1136/bmj.c1066.

[2] Verkooijen HM, Roes K, van Gils CH. Cohort multiple randomized controlled trial: a solution for the evaluation of multiple interventions. Ned Tijdschr Geneeskd 2013;157(17):A5762 ['Cohort multiple randomized controlled trial': een oplossing voor de evaluatie van multipele interventies].

[3] Saesen R, Van Hemelrijck M, Bogaerts J, et al. Defining the role of real-world data in cancer clinical research: the position of the European Organisation for Research and Treatment of Cancer. Eur J Cancer 2023;186:52–61. https://doi.org/10.1016/j.ejca.2023.03.013.

[4] Couwenberg AM, Burbach JPM, May AM, Berbee M, Intven MPW, Verkooijen HM. The trials within cohorts design facilitated efficient patient enrollment and generalizability in oncology setting. J Clin Epidemiol 2020;120:33–9. https://doi.org/10.1016/j.jclinepi.2019.12.015.

[5] Gal R, Monninkhof EM, van Gils CH, et al. The trials within cohorts design faced methodological advantages and disadvantages in the exercise oncology setting. J Clin Epidemiol 2019;113:137–46. https://doi.org/10.1016/j.jclinepi.2019.05.017.

[6] ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf.

[7] Luijken K, van Eekelen R, Gardarsdottir H, Groenwold RHH, van Geloven N. Tell me what you want, what you really really want: Estimands in observational pharmacoepidemiologic comparative effectiveness and safety studies. Pharmacoepidemiol Drug Saf 2023;32(8):863–72. https://doi.org/10.1002/pds.5620.

[8] Ratitch B, Bell J, Mallinckrodt C, et al. Choosing Estimands in clinical trials: putting the ICH E9(R1) into practice. Ther Innov Regul Sci 2020;54(2):324–41. https://doi.org/10.1007/s43441-019-00061-x.

[9] Goetghebeur E, le Cessie S, De Stavola B, Moodie EE, Waernbaum I. On behalf of the topic group causal inference of the Si. Formulating causal questions and principled statistical answers. Stat Med 2020;39(30):4922–48. https://doi.org/10.1002/sim.8741.

[10] Kessels R, May AM, Koopman M, Roes KCB. The trial within cohorts (TwiCs) study design in oncology: experience and methodological reflections. BMC Med Res Methodol 2023;23(1):117. https://doi.org/10.1186/s12874-023-01941-5.
[11] Burbach JP, Kurk SA, Coebergh van den Braak RR, et al. Prospective Dutch colorectal cancer cohort: an infrastructure for long-term observational, prognostic, predictive and (randomized) intervention research. Acta Oncol 2016;55(11): 1273–80. https://doi.org/10.1080/0284186X.2016.1189094.
[12] Couwenberg AM, Burbach MJ, Smits AB, et al. The impact of retractor SPONGE-assisted laparoscopic surgery on duration of hospital stay and postoperative complications in patients with colorectal cancer (SPONGE trial): study protocol for a randomized controlled trial. Trials 2016;17(1):132. https://doi.org/10.1186/s13063-016-1256-x.
[13] Fahim M, Couwenberg A, Verweij ME, Dijksman LM, Verkooijen HM, Smits AB. SPONGE-assisted versus Trendelenburg position surgery in laparoscopic sigmoid and rectal cancer surgery (SPONGE trial): randomized clinical trial. Br J Surg 2022; 109(11):1081–6. https://doi.org/10.1093/bjs/znac249.
[14] Young-Afat DA, Van Gils CH, Van den Bongard H, Verkooijen HM, Group US. The Utrecht cohort for multiple BREast cancer intervention studies and long-term evaLuAtion (UMBRELLA): objectives, design, and baseline results. Breast Cancer Res Treat 2017;164(2):445–50. https://doi.org/10.1007/s10549-017-4242-4.
[15] Gal R, Monninkhof EM, Groenwold RHH, et al. The effects of exercise on the quality of life of patients with breast cancer (the UMBRELLA fit study): study protocol for a randomized controlled trial. Trials 2017;18(1):504. https://doi.org/10.1186/s13063-017-2252-5.
[16] Gal R, Monninkhof EM, van Gils CH, et al. Effects of exercise in breast cancer patients: implications of the trials within cohorts (TwiCs) design in the UMBRELLA fit trial. Breast Cancer Res Treat 2021;190(1):89–101. https://doi.org/10.1007/s10549-021-06363-9.
[17] Bowden J, Bornkamp B, Glimm E, Bretz F. Connecting Instrumental Variable methods for causal inference to the Estimand Framework. Stat Med 2021;40(25): 5605–27. https://doi.org/10.1002/sim.9143.
[18] Lipkovich I, Ratitch B, Qu Y, Zhang X, Shan M, Mallinckrodt C. Using principal stratification in analysis of clinical trials. Stat Med 2022;41(19):3837–77. https://doi.org/10.1002/sim.9439.
[19] Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics 2002;58(1):21–9. https://doi.org/10.1111/j.0006-341x.2002.00021.x.
[20] Stuart EA, Jo B. Assessing the sensitivity of methods for estimating principal causal effects. Stat Methods Med Res 2015;24(6):657–74. https://doi.org/10.1177/0962280211421840.
[21] Olarte Parra C, Daniel RM, Bartlett JW. Hypothetical Estimands in clinical trials: a unification of causal inference and missing data methods. Stat Biopharm Res 2023; 15(2):421–32. https://doi.org/10.1080/19466315.2022.2081599.