ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL ARTICLE

The impact of analysis methodology details on invasive breast cancer-free survival in randomized clinical trials

D. Hlauschek[1], C. Fesl[1] & M. Gnant[1,2]*

[1]ABCSG, Austrian Breast and Colorectal Cancer Study Group, Vienna; [2]Medical University of Vienna, Comprehensive Cancer Center, Vienna, Austria

Available online xxx

Background: Standardized endpoint definitions are crucial for the correct interpretation and reporting of clinical trials. In the field of adjuvant breast cancer clinical trials, the Standardized Definitions for Efficacy End Points (STEEP) criteria were introduced in 2007. In 2021, the STEEP criteria were re-assessed, and a new endpoint was proposed: invasive breast cancer-free survival (IBCFS). However, this recent endpoint also introduces complexity, and detailed considerations of the consequences must be made.

Materials and methods: IBCFS does not include second primary non-breast cancers (SPNBCs). This complicates the statistical analysis as it is no longer perfectly straightforward. We investigate the multiple analysis options and their crucial implications from a theoretical perspective as well as from a practical point of view, using actual trial data from four large adjuvant breast cancer trials.

Results: SPNBCs can either be ignored, censored, or treated as competing events. Importantly, all three approaches address different clinical questions: when SPNBCs are ignored, the 'total' treatment effect is estimated. By censoring SPNBCs, a hypothetical estimand (IBCFS risk had no SPNBCs occurred) is targeted. If SPNBCs are treated as competing events, the IBCFS risk only while subjects remain free from any SPNBC is estimated. In our four large clinical trial dataset example, all three approaches yielded relatively similar results, with the largest differences being observed on the absolute risk scale.

Conclusions: Full standardization of endpoints can only be achieved when the excluded components (like SPNBCs for IBCFS) of the endpoints are considered (and reported in detail!) as well. Different statistical approaches address different clinical questions. For the majority of clinical trials, reporting the total effect ('ignoring' SPNBCs) on IBCFS as well as invasive disease-free survival is strongly recommended.

Key words: endpoints, breast cancer, invasive breast cancer-free survival, second primary non-breast cancers, estimands

INTRODUCTION

Standardized endpoint definitions are crucial in clinical trial reporting which applies in particular to composite endpoints.[1] Invasive disease-free survival (IDFS) is the main endpoint used in contemporary randomized breast cancer trials. According to the Standardized Definitions for Efficacy End Points (STEEP) criteria,[2] this composite endpoint includes ipsilateral invasive breast tumor recurrence, regional invasive breast cancer recurrence, distant recurrence, contralateral invasive breast cancer, death from any cause, and second primary non-breast invasive cancer (SPNBC). In 2021, the STEEP criteria were refined and a 'new' additional endpoint for breast cancer studies was proposed: invasive breast cancer-free survival (IBCFS), which "includes all invasive disease-free survival events except second non-breast primary cancers".[3]

Tolaney et al.[3] base their suggestion of this new endpoint on clinical as well as statistical arguments, the latter being supported by a few simulations. The main 'clinical' argument not to consider other cancer types is 'clinical relevance', and that these 'non-breast primary cancer' events will not be 'informative' with respect to the—usually breast cancer-specific—(therapeutic) question a trial is asking. This, however, can generally be said about other event types as well, e.g. death without prior recurrence, which is a common problem, particularly in elderly trial populations. In any case, the restriction to specific event types included in the primary endpoint raises questions about the statistical analysis method to use, and subsequently, the interpretation of results, particularly in view of their clinical relevance. For early breast cancer trial reporting it is essential that the details of

*Correspondence to: Prof. Michael Gnant, Comprehensive Cancer Center, Medical University of Vienna, Waehringer Gürtel 18-20, A-1090 Vienna, Austria. Tel: +43-1-40400 56460

E-mail: michael.gnant@meduniwien.ac.at (M. Gnant).

endpoint composition are thoroughly conceptualized, transparently reported, and discussed.

In essence, the stipulation of a certain event type as 'less informative' leads to significant interpretation challenges, and this is what also happens with the more recently proposed endpoint IBCFS in the transition from STEEP-1 to STEEP-2. The exclusion of SPNBC from the newly proposed composite endpoint can be handled in three different ways: such events can be either ignored (A), censored (B), or treated as competing event (C) during statistical analysis. Without knowing how SPNBC events were exactly handled, the interpretation of reported IBCFS results is incomplete, and potentially inaccurate. Tolaney et al. reported that "… the estimated HR for IBCFS is insensitive to these events, even when the rates vary between the SOC and experimental arms because of censoring event times for these nonrecurrence events". Hence, they obviously used approach (B), but this warrants a special interpretation and/or the assumption of non-informative censoring. If SPNBCs alter the rate of IBCFS events, which in fact is a rather likely scenario, then the true hazard ratios (HRs) for IBCFS can very well be affected by the SPNBC rates in the two treatment arms. Until recently, this topic was often ignored in the medical literature; however, some reports mention the potential methodological issues with such endpoints (e.g. DATECAN[4]). Furthermore, in 2019, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) consortium published the addendum on estimands[5] which deals with such intercurrent events (events post-baseline that might affect the outcome) and improved the discussion around clinical questions addressed with different analysis strategies.

Since the definition of IBCFS, several retrospective analyses and papers have been published using this endpoint. However, details of dealing with SPNBCs are rarely mentioned and interpretation of results are therefore limited.[6-8] IBCFS is also the primary endpoint in currently ongoing large phase III trials [e.g. CAMBRIA-1 (NCT05774951; EUCT2022-501024-20-00), CAMBRIA-2 (NCT05952557; EUCT 2023-504031-41-00)] which makes the correct understanding of the statistical analysis details of IBCFS even more important.

Therefore, we consider it important that both clinical trialists and study statisticians, but also readership of clinical trial reports, are aware of the consequences of methodology with respect to how exactly event types not included in the defined primary endpoint are handled in clinical breast cancer trials. In this paper, we review the consequences and potential impact of the three mentioned approaches of dealing with 'non-informative' event types. We further demonstrate the importance of this subject using data from four large, randomized, early breast cancer trials.

MATERIALS AND METHODS

IBCFS analysis details

As seen nowadays, for time-to-event result reporting of clinical breast cancer trials such as analysis of 'time to an IBCFS event', a natural starting point is the Kaplan—Meier estimates, accompanied by results from a Cox proportional hazards regression model. However, with endpoints that do not include all event types, the excluded events become intercurrent events,[5,9-11] and one needs to choose how to deal with them. Three potential approaches emerge: they can be either ignored (A), censored (B), or treated as competing events (C). The first two allow the use of the standard time-to-event methods mentioned. However, the last approach requires the competing risk framework (e.g. Fine and Gray sub-distribution hazards regression,[12] cumulative incidence function estimates[13]).

These three analysis options exist every time any event type is not part of the endpoint definition, and not just for IBCFS. For example, when analyzing distant disease-free survival or local-regional recurrence-free survival, SPNBCs but also other event types are not included, and need to be considered during the analysis. For simplicity, this paper focuses only on IBCFS and the handling of SPNBCs, but the main conclusions can be generalized to other endpoints as well.

IBCFS interpretation differences (as compared with IDFS). Maybe not obvious at first, dealing with SPNBC in different ways leads to different risk and hazard estimands that can be best explained using a simple graph as shown in Figure 1A, only considering treatment, SPNBCs, and IBCFS. Treatment may impact IBCFS (Figure 1A, red arrow), which also represents the 'direct' effect of treatment on IBCFS. Furthermore, treatment may influence the SPNBC rate (Figure 1A, left blue arrow) and in turn the occurrence/presence of SPNBC may increase the rate of IBCFS (right blue arrow). These effects visualize the 'indirect' path from treatment to IBCFS via SPNBCs. The 'total' treatment effect would be represented by the 'combination' of the direct and indirect path. Utilizing this visualization, the three analysis options for IBCFS can be summarized as follows:

A. Ignoring all SPNBCs:

By not dealing with SPNBCs, the indirect path (Treatment → SPNBC → IBCFS) stays intact/open and any potential treatment effect via SPNBC on IBCFS is included in the estimate. Furthermore, the estimate includes any potential direct treatment effects (Figure 1A, red arrow) and therefore, the estimated treatment effect represents the total effect of treatment on IBCFS (Figure 1A). Thus—and highly paradoxically—by ignoring SPNBCs they might become 'important' because they end up affecting the composite endpoint.

B. Censoring all SPNBCs:

Censoring implies that the indirect treatment effect path is no longer entirely captured in the analysis as follow-up after an SPNBC is cut short. Hence, all later recurrences are disregarded and do not contribute to IBCFS anymore. In other words, the treatment effect estimate does no longer include any potential effect of treatment on IBCFS via an SPNBC which is visualized in Figure 1B by excluding the

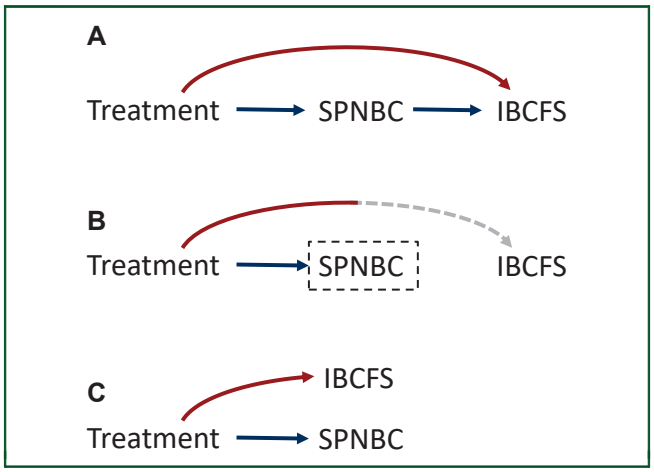

Figure 1. Directed acyclic graph of treatment effects on IBCFS. (A) Simplified visualization of effect paths from treatment to IBCFS in a randomized controlled trial. Red arrow represents the 'direct' treatment effect on IBCFS. Blue arrows represent the 'indirect' treatment effect path via SPNBC. (B) Simplified visualization of remaining effects after censoring for SPNBC. (C) Simplified visualization of the competing event approach. See main text for more information. IBCFS, invasive breast cancer-free survival; SPNBC, second primary non-breast cancer.

second blue arrow. Furthermore, even the direct effect path of treatment on IBCFS may no longer be complete as everything occurring after an SPNBC is censored (represented in the Figure 1B as a dashed line). In simpler terms, let us consider a trial investigating a drug without a constant effect over time; up to the occurrence of an SPNBC, assume the IBCFS rate in the experimental arm is only half the rate in the control arm (HR = 0.5; green solid line in Figure 1B), but somehow the drug effect vanishes afterwards (HR = 1; dashed grey line). The entire direct treatment effect in this example is clearly something between 0.5 and 1, but as all follow-up is censored after an SPNBC event, the estimate would be 0.5. In other words, the censoring approach assumed that the direct effect on IBCFS before and after an SPNBC is identical. Thus, the estimated effect is not the total effect and—in worst case—only reflects part of the direct treatment effect. Standard statistical methods (e.g. Cox model, Kaplan—Meier estimates) assume non-informative censoring, which implies that patients experiencing an SPNBC have a similar IBCFS rate as patients who did not experience an SPNBC. In reality, this assumption is not very realistic: an SPNBC event very likely increases the rate of death, which is one contributor of the IBCFS endpoint. Therefore, the censoring strategy aims at a different 'target' compared with approach A. In fact, the results from such an analysis need to be interpreted from a hypothetical point of view where SPNBCs are eliminated.

C. Treating all SPNBCs as competing events:

This again implies that all later events are not considered during analysis. But, because SPNBCs are 'considered' terminal events in standard competing event analysis (i.e. no further events can occur), this is different from approach B. In fact, it means we are kind of assuming an IBCFS event rate of 0% after the occurrence of SPNBCs. Therefore, in Figure 1C, no direct path from SPNBC to IBCFS is shown. Furthermore, and different from the censoring approach, we are no longer assuming the direct treatment effect stays the same before and after an SPNBC. With the competing event approach, we are only interested in what happens before an SPNBC (visualized in Figure 1C by alignment of IBCFS above SPNBC). Therefore, it follows that the treatment effect estimates refer to the hazard/risk before an SPNBC. Of note, this is often referred to as 'semi-competing' risk setting, because SPNBCs are no immediate 'terminal' events like death in 'standard' competing risk settings (see Fine et al.[14] for example).

A brief summary of the three approaches can be found in Table 1. It needs to be noted that Figure 1A-C are simplifications only applicable to controlled randomized trials. We intentionally did not include more complex versions with two or more time points, to make the interpretation easier. More in-depth reviews of total and direct effects in a standard competing risk setting are available.[15,16]

Theoretical implications

Based on the above-mentioned considerations, the following can be deduced:

- Approach A will result in the largest number of events, and approaches B and C will have the same number of events.
- All three approaches can lead to different 'absolute' cancer risk estimates. Approach C will always yield the lowest risk estimates as it is assumed that no further IBCFS events can occur after an SPNBC. Approaches A and B will only show the same results (on average) if the censoring of SPNBCs is truly non-informative. Often, approach A will show the highest absolute risk because patients with SPNBCs will have a higher risk of death compared with event-free patients in clinical reality.
- The actual magnitude of difference between the three methods strongly depends on the underlying SPNBC rate in the population. The smaller the rate in comparison to the IBCFS rate, the lesser the impact of censoring or treating it as a competing event. So, one may expect larger differences in cohorts with overall lower breast cancer risk.
- Approach A requires complete follow-up of all patients until the IBCFS event. If, however, follow-up after an SPNBC is no longer required per protocol, the estimation may be biased and only approaches B and C can be used.
- Most importantly, because all three statistical approaches handle those intercurrent events differently, they address different estimands/clinical objectives (Table 1)[5]:
  - Does treatment improve IBCFS (irrespective of the occurrence of SPNBCs)?
  - Does treatment improve IBCFS in the absence of SPNBCs?
  - Does treatment improve IBCFS while being free of any SPNBC?

Table 1. Summary of three approaches and their effect estimate interpretation when dealing with SPNBCs in time to IBCFS analysis

Handling of SPNBC in the analysis	Treatment effect estimand and interpretation	Related trial objective
Ignoring	IBCFS risk after including all direct and indirect effects of treatment (real world)	Does treatment improve time to IBCFS (regardless of SPNBCs occurrence)?
Censoring	IBCFS risk under elimination (=censoring) of SPNBC (hypothetical world)	Does treatment improve time to IBCFS had no SPNBCs occurred?
Competing event	IBCFS risk before an SPNBC	Does treatment improve time to IBCFS while being free of SPNBCs?

IBCFS, invasive breast cancer-free survival; SPNBC, second primary non-breast cancer.

Notably, since the ICH addendum on estimands, those three approaches are often referred to as (i) treatment policy strategy, (ii) hypothetical strategy, and (iii) while-on-treatment strategy.[5,9,10]

Two further (more technical) implications—when there is a treatment effect on SPNBC's—when using those three approaches can be found in the Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2025.105324.

To summarize the theoretical considerations about dealing with SPNBCs: neither approach addresses all scientific questions that might be of interest. However, approach A might be preferred for two reasons: firstly, because of its methodological simplicity (i.e. SPNBCs need not be handled in any way), and secondly, because it is the only method that provides the total treatment effect that is linked to the natural question that first comes to mind when confronted with cancer—if I take this drug, does it decrease my risk of a recurrence or death (i.e. an IBCFS event)? The interpretation of results using approaches B and C require more reflection and may not be of 'immediate' interest to a patient/clinician when discussing treatment options.

Statistical analysis

To further investigate the three theoretical means of dealing with SPNBC and its consequences on reported trial results, we have tested them in four large clinical breast cancer trials: Cox models and Fine and Gray sub-distribution hazards models, as well as Kaplan–Meier and cumulative incidence estimates were applied. Recurrences and SPNBC events with the same date were treated as BCFS events. Results include (sub-distribution) HRs and their respective 95% confidence intervals (CIs) and *P* values. The proportional hazards assumption was assessed with an interaction term with time added to the regression models. Time points for absolute risk estimates were chosen based on clinical meaningfulness. All analyses were carried out with SAS 9.4.

RESULTS

The results are from large prospective randomized early breast cancer trials as examples for endpoint definition strategies. All four trials, conducted by the Austrian Breast and Colorectal Cancer Study Group (ABCSG), investigated a new treatment (combination) in a randomized, parallel, two-arm setting (only ABCSG-12 was a 2 × 2 factorial design). A summary of these can be found in Table 2. All four trials investigated (time to) DFS as the primary endpoint. Further details of these studies can be found in the main trial publications.[17-24] It should be noted that as these older trials did not distinguish between invasive and non-invasive disease, all endpoints are 'reduced' to BCFS and DFS.

The number of patients included in these analyses ranged from 1803 (ABCSG-12) up to 3901 (ABCSG-8). Frequency of SPNBCs was around 6%-7%, except for ABCSG-12 where it was only 2.2% (Table 2), due to much younger patient age in this study. The total number of DFS events (Table 3) ranged from 264 (14.6%, ABCSG-12) to 926 (23.7%, ABCSG-8).

Table 3 shows BCFS results of all three approaches, as well as DFS results for completeness. All four trials show the same pattern for absolute risk: expectedly, the estimated BCFS risk is always largest for approach A, followed by the censoring and competing event approaches. The differences between those methods range from <1% (ABCSG-12) up to 3.5% (ABCSG-18). As an example, the differences in estimated BCFS risk over the entire follow-up is shown for ABCSG-8 (Figure 2).

In terms of absolute treatment effect sizes, the different analysis strategies yielded rather similar estimates, except for ABCSG-18 where a risk difference between denosumab and placebo of 3.3% was estimated with approach A, and only a 2.0% difference was estimated with the competing event method B (Table 2D).

Relative treatment effects measured with HRs were similar across the three approaches, except for ABCSG-18 where a HR of 0.80 (95% CI 0.67-0.96) was estimated with approach A, and slightly weaker effects using the other approaches (B: HR = 0.86, 95% CI 0.70-1.04; C: HR = 0.87, 95% CI 0.71-1.05). Remarkably, if this had been a trial with BCFS as primary endpoint, only approach A would have yielded a positive study ($P = 0.0146$). The trial would have missed its target if analyzed primarily with approach B or C ($P = 0.1152$; $P = 0.1499$).

Comparing the BCFS with the DFS results, the 'relative' treatment effects are surprisingly similar, with approach A yielding the closest results. However, the 'absolute' treatment effects between BCFS and DFS vary in a non-neglectable way. For example, in ABCSG-18, the absolute BCFS difference was only 2.0% according to the competing event approach but more than doubled for DFS (4.5%). It is also noteworthy that the largest absolute treatment effects (1.1% up to 4.5%) are only seen with the DFS endpoint.

The occurrence of SPNBC was very similar across treatment arms, except for ABCSG-18, where less SPNBCs occurred in the active therapy group arm. Results for ABCSG-18 and three different analysis strategies for SPNBCs can be found in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105324. Only minor

Table 2. Overview of adjuvant early breast cancer trials: ABCSG-6, ABCSG-8, ABCSG-12, and ABCSG-18

Trial	Treatment regime	Population	*N*	Frequency of SPNBC
ABCSG-6	2 years of tamoxifen + aminoglutethimide followed by 3 years of tamoxifen versus 5 years of tamoxifen alone	Postmenopausal patients with stage I and II, hormone receptor-positive breast cancer	2020	143 (7.1%)
ABCSG-8	5 years of tamoxifen versus 2 years of tamoxifen followed by 3 years anastrozole	Postmenopausal patients with hormone receptor-positive, G1 or G2 breast cancer	3901	264 (6.8%)
ABCSG-12	Anastrozole versus tamoxifen and both, with or without zoledronic acid for 3 years	Premenopausal patients with endocrine-responsive early breast cancer and ovarian function suppression	1803	39 (2.2%)
ABCSG-18	Denosumab versus placebo every 6 months until aromatase inhibitor discontinuation	Postmenopausal patients with early, hormone receptor-positive breast cancer on adjuvant aromatase inhibitor therapy	3420	259 (7.6%)

As these older trials did not distinguish between invasive and non-invasive disease, all endpoints are 'reduced' to BCFS and DFS.
BCFS, breast cancer-free survival; DFS, disease-free survival; SPNBC, second primary non-breast cancer.

differences can be seen with respect to the absolute and relative effect sizes.

DISCUSSION

Different estimands can be derived when using IBCFS as clinical trial endpoint, depending on the details of how to handle SPNBCs. We confirm that the full event breakdown is very important to judge the treatment effect,[1] but in addition demonstrate here that the three analysis methods described address different scientific questions.

The competing event approach (C) might be of minor clinical interest as it describes only the treatment effect before an SPNBC. In addition, the potential for drawing wrong conclusions is higher than for the other two methods: a positive drug effect on SPNBCs can translate into a worse treatment effect for IBCFS, because a positive drug effect means fewer SPNBC events and therefore more opportunity for an IBCFS event in the experimental group.

The censoring approach (B) might be more useful but still needs to be interpreted very carefully as it estimates the drug effect in a hypothetical world where no SPNBCs occur. If one is willing to assume that the direct treatment effect on IBCFS does not vary over time, approach B might still be useful in estimating the direct effect of treatment on IBCFS regardless of any indirect treatment effects. As a side note, there are methods that try to adjust for non-informative censoring like, for example, inverse probability censoring weighting.[25] However, as they only adjust for everything measured 'before' the censoring time point, they may not compensate for a changing treatment effect 'after' the censoring. Another drawback of approach B is that the estimates can never be verified since it is virtually impossible to eliminate SPNBCs, even in a perfectly conducted clinical trial: the real world always reflects a composition of direct and indirect treatment effects.

Only with approach (A) one can estimate those total treatment effects. Theoretically, it can also be argued that even if there is a positive direct treatment effect, the total effect is still worse in the treatment group, because of potential detrimental indirect effects, like when treatment increases the SPNBC rates in a non-neglectable manner. Hence, only the total treatment effect estimated ignoring SPNBCs bears the lowest risk of misinterpretation and is most advisable. Furthermore, approach (A) reflects a potential negative effect of treatment on the SPNBC event rate 'correctly', as the IBCFS risk increases in the treatment arm.

While we propose that these considerations should lead to obligatory reporting of how 'non-relevant' event types have been handled in any clinical trial report, the practical examples show that the differences between the three methods are often minor—regardless of the correct interpretation of the statistical estimands. However, as seen with ABCSG-18, the example in which also the greatest association between treatment arms and SPNBC rates was observed, the estimates and trial conclusions might end up differently. In the worst case, choosing an inappropriate statistical method could address the wrong scientific question and result in a trial that is declared negative although there actually exists a (total) effect of the therapeutic intervention studied.

Where does this leave us in terms of planning and conducting clinical trials with IBCFS? As demonstrated above, the total effect should be considered first when choosing the primary analysis strategy: clinical relevance should be the main decision driver for choosing the main estimand. When the main goal is the implementation of a new treatment (regime), the total effect is of greatest importance as it probably closest reflects the real world.

Nevertheless, if it is of interest to assess the potential treatment effect when there are no SPNBCs or the effect before any SPNBC, it also might be worthwhile investigating the effects estimated using censoring and competing event methodology. It can be argued that analysis of IDFS should also be investigated, as only this gives full insights; both IBCFS and IDFS should be reported to the scientific community and regulators. In addition, analysis of time to SPNBC (Caveat: this can be analyzed in several different ways, too) can yield additional valuable insights and should routinely be done in settings with higher second cancer prevalence (i.e. higher patient age).

In today's—fortunately—innovation-rich world of clinical breast cancer trials, accurate reporting is crucially important.

Table 3. BCFS and DFS results from four large, randomized, clinical trials: ABCSG-6, ABCSG-8, ABCSG-12, and ABCSG-18

					Absolute risk		Cox regression	
Endpoint	Arm	*N*	Events	Competing	10-year	Diff.	HR[a] (95% CI)	*P* value
ABCSG-6								
BCFS (ignored)	TAM + AG	1012	369	NA	31.6%	−1.3%	0.94 (0.82-1.08)	0.3938
	TAM only	1008	390	NA	32.9%		1	
BCFS (censored)	TAM + AG	1012	345	NA	30.3%	−1.6%	0.95 (0.82-1.11)	0.5302
	TAM only	1008	358	NA	31.9%		1	
BCFS (competing)	TAM + AG	1012	345	63	29.4%	−1.4%	0.96 (0.83-1.11)	0.5855
	TAM only	1008	358	70	30.8%		1	
DFS	TAM + AG	1012	408	NA	34.8%	−1.7%	0.94 (0.82-1.08)	0.3975
	TAM only	1008	428	NA	36.5%		1	
ABCSG-8								
BCFS (ignored)	2-year TAM + 3-year ANA	1946	377	NA	23.3%	−1.0%	0.93 (0.80-1.07)	0.2806
	5-year TAM	1955	399	NA	24.3%		1	
BCFS (censored)	2-year TAM + 3-year ANA	1946	327	NA	21.6%	−0.6%	0.93 (0.80-1.08)	0.3167
	5-year TAM	1955	344	NA	22.3%		1	
BCFS (competing)	2-year TAM + 3-year ANA	1946	327	130	20.6%	−0.6%	0.93 (0.80-1.08)	0.3157
	5-year TAM	1955	344	125	21.2%		1	
DFS	2-year TAM + 3-year ANA	1946	457	NA	28.4%	−1.1%	0.95 (0.84-1.08)	0.4341
	5-year TAM	1955	469	NA	29.4%		1	
ABCSG-12								
BCFS (ignored)	Zoledronic acid	900	107	NA	12.8%	−3.4%	0.79 (0.61-1.02)	0.0709
	Control	903	130	NA	16.3%		1	
BCFS (censored)	Zoledronic acid	900	102	NA	12.3%	−3.6%	0.78 (0.60-1.01)	0.0563
	Control	903	126	NA	15.9%		1	
BCFS (competing)	Zoledronic acid	900	102	17	12.2%	−3.5%	0.78 (0.60-1.01)	0.0580
	Control	903	126	19	15.7%		1	
DFS	Zoledronic acid	900	119	NA	14.2%	−3.9%	0.79 (0.62-1.00)	0.0521
	Control	903	145	NA	18.1%		1	
ABCSG-18								
BCFS (ignored)	Denosumab	1711	211	NA	17.1%	−3.3%	0.80 (0.67-0.96)	0.0146
	Placebo	1709	264	NA	20.4%		1	
BCFS (censored)	Denosumab	1711	189	NA	15.7%	−2.3%	0.86 (0.70-1.04)	0.1152
	Placebo	1709	219	NA	18.1%		1	
BCFS (competing)	Denosumab	1711	189	108	15.0%	−2.0%	0.87 (0.71-1.05)	0.1499
	Placebo	1709	219	141	16.9%		1	
DFS	Denosumab	1711	297	NA	22.9%	−4.5%	0.82 (0.70-0.95)	0.0101
	Placebo	1709	360	NA	27.4%		1	

SPNBCs were either ignored, censored, or treated as competing events in BCFS analyses. Side note: As these older trials did not distinguish between invasive and non-invasive disease, all endpoints are 'reduced' to BCFS and DFS.

BCFS, breast cancer-free survival; CI, confidence interval; DFS, disease-free survival; Diff., Difference; HR, hazard ratio; NA, not applicable; SPNBC, second primary non-breast cancer; TAM, tamoxifen; AG, aminoglutethimide; ANA, anastrozole.

[a]For competing risk analysis, the HR reflects the sub-distribution HR.

Billions of monetary investments, regulatory approvals, and patient treatments depend on clinical trials that either report a difference or not. Clinical trial methodology, together with meticulous rules on data collection and trial governance, aims at minimizing the risk of confounders and biases. Thus, full methodological transparency is mandatory for the accurate interpretation of a clinical trial result. Even with the utmost transparency and correctness, we are only approaching reality with our research results that remain estimates of true effects; some remaining uncertainties cannot be resolved: e.g. when a breast cancer patient develops masses in her lungs or bones, disease recurrence is assumed (and an event recorded)—usually there is no mandatory histologic confirmation, and it eventually remains uncertain whether metastases from breast cancer (which would correctly be counted as IBCFS events) or a primary lung/bone malignancy (which would be a—potentially deadly—SPNBC) are the underlying truth for that radiological finding.

We acknowledge that limitations exist of what can be discussed in this paper: Firstly, only the simplest clinical research setting—controlled randomized trials—has been investigated here. The analysis choice and the estimated effects from observational data or more complex trial designs might be different, and most likely even more complex. For example, creating confounding in observational data due to censoring SPNBCs is a valid concern (new paths in Figure 1 would be created). Secondly, the important issue of the impact of misclassification of SPNBCs (or distant recurrences, as mentioned above) on composite endpoints must be discussed separately. In theory, and most likely often in breast cancer trials, this can also have an impact on the estimates. One recommended option is to check the robustness of the results with sensitivity analyses.[26] When a high misclassification rate is expected, analysis of IBCFS might not be considered at all, and IDFS should remain the primary choice. Thirdly, investigated effect sizes in the

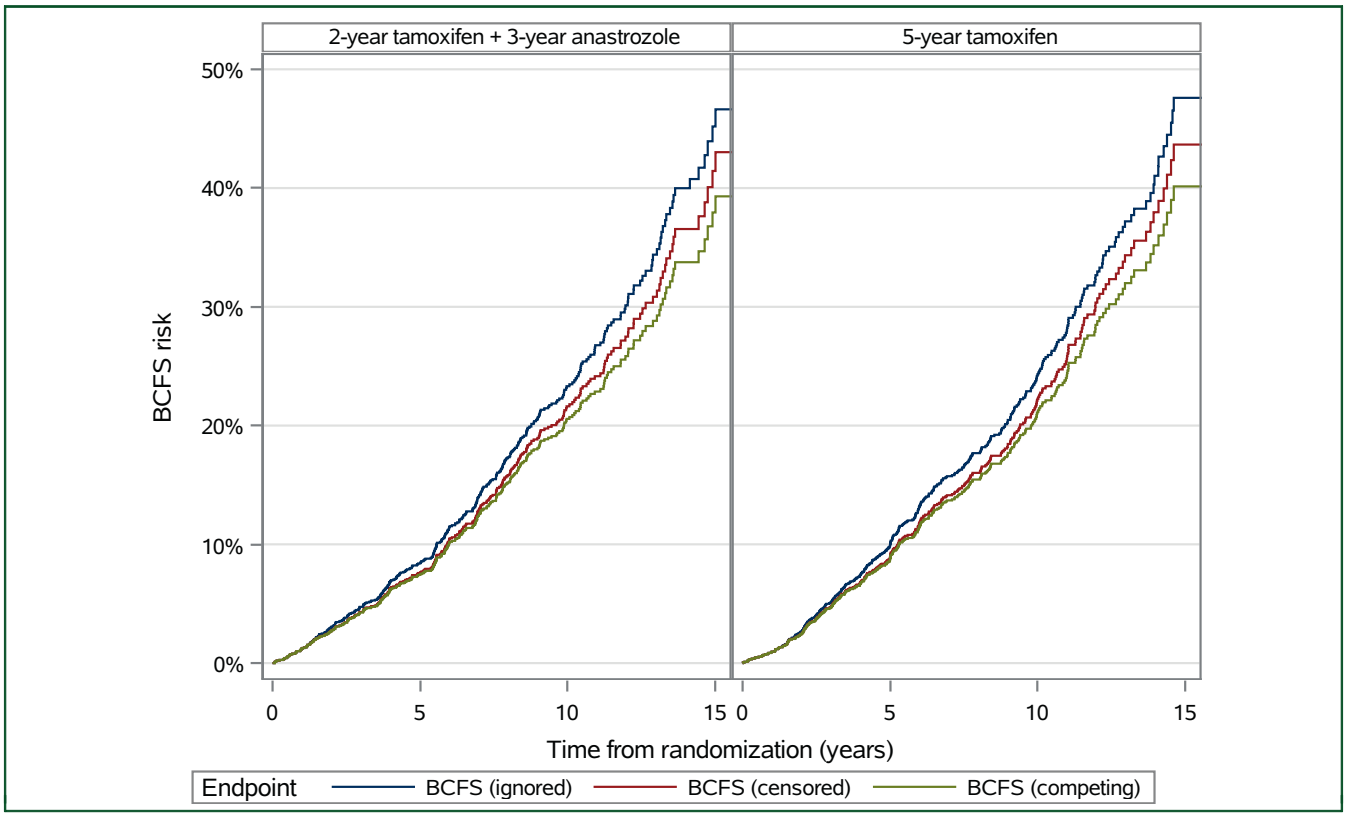

Figure 2. BCFS risk in the ABCSG-8 trial arms according to three different statistical approaches. BCFS risk estimated when ignoring (blue line), censoring (red line), or treating SPNBCs as competing events (green line).
BCFS, breast cancer-free survival; SPNBC, second primary non-breast cancer.

clinical trial examples shown here are modest at best. Trials with large treatment effects might yield greater differences, and further analysis of such studies could be useful.

Conclusions

In summary, defining and clearly reporting the estimand/objective is of utmost importance to avoid misinterpretation of clinical trial results. Most often, reporting the total effect on IBCFS as well as IDFS might be the best choice and it is recommended to not miss important negative effects of treatment on SPNBCs. Future guidelines should reflect the challenges with 'incomplete' endpoints, i.e. endpoints that do not include all post-randomization event types. As demonstrated in this paper, even a 'harmonized' endpoint definition like IBCFS is ultimately not sufficient to truly harmonize clinical trial result reporting. It is not sufficient to describe only the event types to be included in trial endpoints. Reporting guidelines both by regulatory agencies as well as scientific journals should make the detailed reporting of exact endpoint assessment methodology mandatory.

FUNDING

None declared.

DISCLOSURE

MG reports personal fees/travel support from Amgen, AstraZeneca, Bayer, Daiichi Sankyo, Eli Lilly, EPG Health (IQVIA), Menarini-Stemline, MSD, Novartis, Pierre Fabre, and Veracyte. All other authors have declared no conflicts of interest.

REFERENCES

1. Walia A, Tuia J, Prasad V. Progression-free survival, disease-free survival and other composite end points in oncology: improved reporting is needed. *Nat Rev Clin Oncol*. 2023;20:885-895.
2. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*. 2007;26:2127-2132.
3. Tolaney SM, Garrett-Mayer E, White J, et al. Updated standardized definitions for efficacy end points (STEEP) in adjuvant breast cancer clinical trials: STEEP version 2.0. *J Clin Oncol*. 2021;39:2720-2731.
4. Gourgou-Bourgade S, Cameron D, Poortmans P, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol*. 2015;26:873-879.
5. ICH. Harmonised Guideline. *Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1)*. International Council for Harmonisation of Technical Requirements for Pharmaceuticals For Human Use; 2017.
6. Tesch ME, Zheng Y, Rosenberg SM, et al. Estrogen levels in young women with hormone receptor-positive breast cancer on ovarian function suppression therapy. *NPJ Breast Cancer*. 2024;10:67.
7. Wang Y, Tesch ME, Lim C, et al. Risk of recurrence and pregnancy outcomes in young women with breast cancer who do and do not undergo fertility preservation. *Breast Cancer Res Treat*. 2022;195:201-208.
8. Blondeaux E, Xie W, Carmisciano L, et al. Intermediate clinical endpoints in early-stage breast cancer: an analysis of individual patient

data from the Gruppo Italiano Mammella and Mammella Intergruppo trials. *EClinicalMedicine*. 2024;70:102501.
9. Casey M, Degtyarev E, Lechuga MJ, et al. Estimand framework: are we asking the right questions? A case study in the solid tumor setting. *Pharm Stat*. 2021;20:324-334.
10. Rufibach K. Treatment effect quantification for time-to-event endpoints-Estimands, analysis strategies, and beyond. *Pharm Stat*. 2019;18:145-165.
11. Li R, Zhang J, Wang J, Wang J. Statistical considerations in long-term efficacy evaluation of anti-cancer therapies. *Front Pharmacol*. 2023;14:1265953.
12. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496-509.
13. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133:601-609.
14. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001;88:907-919.
15. Rudolph JE, Lesko CR, Naimi AI. Causal inference in the face of competing events. *Curr Epidemiol Rep*. 2020;7:125-131.
16. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, et al. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med*. 2020;39:1199-1236.
17. Gnant M, Mlineritsch B, Schippinger W, et al. Endocrine therapy plus zoledronic acid in premenopausal breast cancer. *N Engl J Med*. 2009;360:679-691.
18. Gnant M, Mlineritsch B, Stoeger H, et al. Adjuvant endocrine therapy plus zoledronic acid in premenopausal women with early-stage breast cancer: 62-month follow-up from the ABCSG-12 randomised trial. *Lancet Oncol*. 2011;12:631-641.
19. Gnant M, Mlineritsch B, Stoeger H, et al. Zoledronic acid combined with adjuvant endocrine therapy of tamoxifen versus anastrozol plus ovarian function suppression in premenopausal early breast cancer: final analysis of the Austrian Breast and Colorectal Cancer Study Group Trial 12. *Ann Oncol*. 2015;26:313-320.
20. Gnant M, Mlineritsch B, Luschin-Ebengreuth G, et al. Adjuvant endocrine therapy plus zoledronic acid in premenopausal women with early-stage breast cancer: 5-year follow-up of the ABCSG-12 bone-mineral density substudy. *Lancet Oncol*. 2008;9:840-849.
21. Jakesz R, Jonat W, Gnant M, et al. Switching of postmenopausal women with endocrine-responsive early breast cancer to anastrozole after 2 years' adjuvant tamoxifen: combined results of ABCSG trial 8 and ARNO 95 trial. *Lancet*. 2005;366:455-462.
22. Schmid M, Jakesz R, Samonigg H, et al. Randomized trial of tamoxifen versus tamoxifen plus aminoglutethimide as adjuvant treatment in postmenopausal breast cancer patients with hormone receptor-positive disease: Austrian breast and colorectal cancer study group trial 6. *J Clin Oncol*. 2003;21:984-990.
23. Gnant M, Pfeiler G, Dubsky PC, et al. Adjuvant denosumab in breast cancer (ABCSG-18): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet*. 2015;386:433-443.
24. Dubsky PC, Jakesz R, Mlineritsch B, et al. Tamoxifen and anastrozole as a sequencing strategy: a randomized controlled trial in postmenopausal patients with endocrine-responsive early breast cancer from the Austrian Breast and Colorectal Cancer Study Group. *J Clin Oncol*. 2012;30:722-728.
25. Chesnaye NC, Stel VS, Tripepi G, et al. An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J*. 2021;15:14-20.
26. Bakoyannis G, Yiannoutsos CT. Impact of and correction for outcome misclassification in cumulative incidence estimation. *PLoS One*. 2015;10:e0137454.