SUST J Sci Tech, Vol 34(2) 2024,1-24 ISSN 1682-895X

A Review of Methods for Estimating Treatment Effects Using Single-Arm Studies in NICE Technology Appraisals

Nahid Sultana

Department of Statistics, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh

*Corresponding Author's Email: nahid-sta@sust.edu

Abstract

OBJECTIVES: The aim of this review was to find the techniques utilised in National Institute for Health and Care Excellence (NICE) single technology appraisals (STAs) to assess relative treatment effects with single-arm studies. Evaluating the suitability of the adjustment strategies employed by the indicated approaches is crucial.

METHODS: All the STAs identified within 1st January 2018 to 31 December 2023 in NICE website have been searched to identify relevant STAs involving single-arm studies. Moreover, information was extracted on how prognostic and effect modifier variables have been identified and how survival extrapolation has been conducted in a data extraction form.

RESULTS: Sixteen (76%) of the 20 oncology TAs that participated in single-arm pivotal studies employed population adjustment techniques, primarily Matching Adjusted Indirect Comparison (MAIC) (62%) and Simulated Treatment Comparison (STC) (19%). More than half (57%) of the TAs used multiple treatment comparisons, frequently going beyond the paired goal of MAIC/STC without satisfying the "shared effect modifier" criterion. Study independence was compromised when intervention IPD was used multiple times in a larger disconnected network. Parametric models fitted to unadjusted Individual patient data (IPD) were commonly employed for extrapolation, and MAIC/STC-adjusted HRs were utilised to determine comparator survival. Transportability of the treatment effect was postulated but never supported when directly extrapolated from uncorrected K-M curves.

CONCLUSIONS: It is not surprising that adjustment methods like unanchored MAIC and STC were frequently used in NICE STAs with the increase of single-arm studies. These methods make assumptions about the variables included in the analysis which are difficult to satisfy and can produce residual bias.

Keywords: Single technology appraisals (STA); Single-arm study; Matching Adjusted Indirect Comparison (MAIC); Simulated treatment comparison (STC); Individual patient data (IPD).

1. Introduction

The Health Technology Assessment (HTA) methodically assesses the advantages and disadvantages of health technology in relation to current alternatives. Randomised controlled trials (RCTs), the gold standard for reducing bias, are the best way to examine the effectiveness of pharmaceutical treatments. RCTs, however, are expensive, time-consuming, and frequently lack direct comparisons with rival treatments [2].

Indirect treatment comparisons (ITCs) are employed when head-to-head evidence is not available through RCT. Indirect comparison is in anchored form when the comparison is conducted via common comparators as in diagram (a), (b) in Figure 1. An unanchored indirect comparison takes place when treatments are compared across studies without considering randomisation within studies for the lack of a common comparator as in diagram (d), (e) in Figure 1.

A significant source of bias in any ITC is imbalances in prognostic factors and effect modifiers between study populations, which undermine the "constancy of relative effects" and violate the "constancy of absolute effects" assumptions [6]. Since these imbalances are unavoidable, no recognised ITC approach depends on the "constancy of absolute effects." These imbalances must be taken into account by both anchored and unanchored ITCs. This can

https://doi.org/10.63512/sjst.2024.2001
Received May 26, 2025, Accepted February 3, 2026.

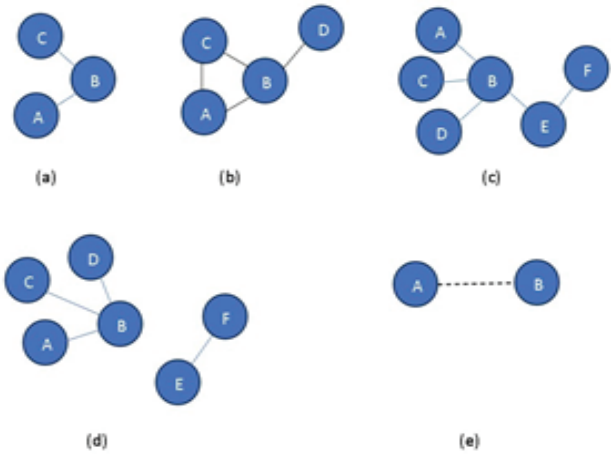

Figure 1: Different forms of a network in indirect comparison

be accomplished by population adjustment using individual patient-level data (IPD). Using "conditional constancy of relative effects" for anchored ITCs and "conditional constancy of absolute effects" for unanchored ITCs, these population adjustment methods minimises discrepancies between study groups and estimates treatment effects in a single target population.

Single-arm studies are a special case of an unanchored comparison comprising a prospective series of patients receiving the same treatment. When spontaneous patient recovery is implausible, placebo effects are negligible, or placebo assignment is unethical, such studies are employed. Unlike RCTs, single-arm studies do not have a comparator group, hence indirect comparison is the main method for estimating relative treatment effects. Therefore, the use of single-arm studies involves reference to an external comparator. Using an external comparator to estimate a treatment effect can induce several limitations. First, the choice of the comparator is often post-hoc, so the probability that it was intentionally selected to favor the new treatment is non-trivial; second, the use of an external comparator can suffer from confusion bias which occurs as no assurance can be given that compared groups are comparable in terms of patients [1]

NICE, a public body of the Department of Health and Social Care in England, provides guidance for health and social care practitioners. Single technology appraisals (STAs) evaluate a treatment for a specific indication, comparing it directly or indirectly with standard care using evidence on benefits, harms, and costs. Appraisals include committee papers, a consultation document (ACD) with provisional recommendations, and a final appraisal document (FAD). These documents assess clinical effectiveness and cost-effectiveness, including the incremental benefit per quality-adjusted life year (QALY) compared to relevant comparators.

In HTA, when companies analyse their intervention treatment that comes from a single-arm study with comparator/comparators, access to IPD in all studies of interest is a rare situation as sharing of clinical data is often limited. A middle-ground situation is more realistic where the company has access to IPD for its own study and aggregate data (AgD) for the comparator studies.

A systematic review conducted by Phillippo et al. (2019b) characterised the use of population adjustment methods in Technical Appraisal (TA)s submitted to NICE from 2010 to 2018 as well as provide recommendations on how and when population adjustment methods should be used. They focused on how these methods were implemented for any outcome in both anchored and unanchored situations without restricting their search for single-arm studies. In Phillippo et al. (2019a) review, 7% (18/268) of the TAs found to be using population adjustment methods.

Though the search was not restricted to unanchored cases most of the population adjustment methods were found for unanchored comparison (89%, 16/18). Eighty-three percent (15/18) of the application found out for oncology where 89% (16/18) uses MAIC and 17% (3/18) uses STC. The criteria that were used for the inclusion of covariates in the adjustment were effective sample size, expert opinion, availability, cross-validation, or statistical significance. Fifty-six percent (10/18) of the comparisons were conducted for a larger network of evidence where comparisons were made for multiple comparators and/or multiple aggregate study populations.

The current review has focused only on unanchored comparisons with the single-arm study. The time frame of this review is 2018 to 2023. This time frame has been chosen to include more recent STAs information. In addition to the information extracted by Phillippo et al. (2019b) [7], this review has extracted information on additional issues including how prognostic and effect modifier variables have been identified and how survival extrapolations have been conducted.

The aim of this review of NICE STAs is to determine how comparisons against relevant comparators have been performed using evidence about new treatments obtained from single-arm studies when IPD is available partially. Additionally, this review aims to evaluate the suitability of the adjustment strategies used by the identified approaches and to assess whether the recommended practices suggested by Phillippo et al. (2019b) have been followed or not.

Section 2 of this paper narrates how the review was conducted and Section 3 describes the results. The paper concludes with a discussion of the results in Section 4 followed by a conclusion in section 5.

2. Methods

2.1 Inclusion criteria

STAs based on single-arm study and published on the NICE website from 1st January 2018 to 31st December 2023 have been included in this review. Although the search do not cover multiple databases, data synthesis (meta-analysis), and risk of bias across studies: it has clear question and rationale (PRISMA Checklist Items 3 & 4), inclusion/exclusion criteria (applicability), information sources, search strategy and data collection process (PRISMA Checklist Items 5, 7, & 9).

2.2 Data extraction

All STAs listed on the NICE website (https://www.nice.org.uk/guidance/published) from 2018 to 2023 have been screened one by one to identify relevant STAs involving single-arm studies. The review has excluded appraisals that had IPD from all studies and included those with partial IPD. This review has extracted information from relevant documents such as committee papers, ACD and FAD. Committee papers are the main document that has been used for data extraction. If committee papers are not available then ACD and FAD have been used for screening purposes.

A data extraction spreadsheet was created using Microsoft Excel which was used to collect relevant information about the methods that have been used to make an indirect comparison between different comparators. Only one reviewer was involved in screening and extraction, no inter-rater agreement was assessed using single-arm studies. Information has been extracted regarding various issues. The question that were asked during the review is summarised in Appendix table 1. Furthermore, the extracted information in complied in Appendix table **A.1- A.7.**

3. Results

A total of 260 TAs have been found which have been published from January 2018 to December 2023. Of these identified TAs, 27 appraisals were identified where the pivotal study/studies were single-arm studies. Of these 27 identified TAs, seven of these had IPD available from all included studies, so they were excluded and 20 TAs were identified with only partial IPD. The PRISMA diagram in Figure 2 demonstrates the selection process.

3.1 Clinical area and outcome of the published TA

All the TAs included in this review since 2018 have been in oncology (20 out of 20, 100%). Other than oncology, no other clinical area has been identified which shows a frequent use of single-arm study.

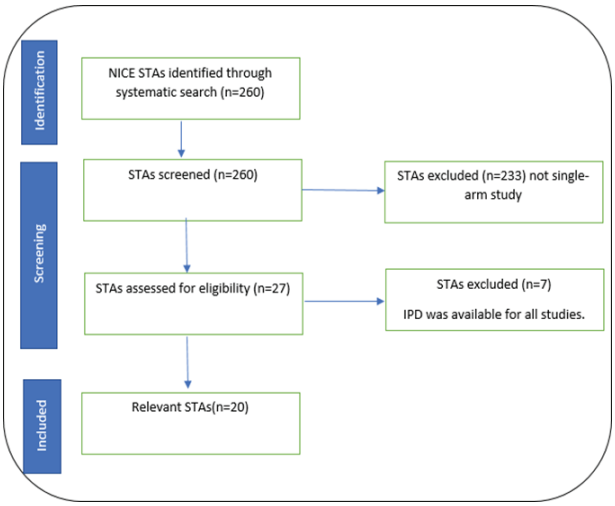

Figure 2: PRISMA diagram for inclusion of relevant NICE STAs with single-arm study

Of these TAs, 16 (80%) have used population adjustment methods and four TAs did not make any kind of adjustment. As all the TAs were on oncology, the outcomes that were indirectly compared were time- to-event outcomes. In all the included TAs, progression-free survival (PFS) and overall survival (OS) were the most common outcome types used in population-adjusted analyses.

3.2 Population Adjustment method

Matching adjusted indirect comparison (MAIC) was mostly used population adjustment methods (13 out of 20, 65%). Simulated treatment comparison (STC) was also used as a population adjustment method, but the frequency of using STC was not as MAIC. STC was used in 4 out of 20 appraisals (20%). Two appraisals used both MAIC and STC [19, 29]. The application of different methods has been summarised in Table 1.

Of these 13 TAs which have used MAIC, only 5 [10, 18, 19, 20, 29] (38.46%) of them have reported their effective sample size (ESS) and none of the TAs mentioned whether the propensity score model for MAIC

Table 1: Application of different comparison methods

Applied method/methods	Number of appraisals
Only MAIC	12
Only STC	3
Both MAIC and STC	1
No adjustment	4
Total	20

includes second-order terms or not. Of these, the median ESS was 67.1 (range: 3.8 to 84), with a median reduction in ESS from the original sample size of 50.3% (range: 43.24% to 94.73%). A huge decrease in ESS indicates lack of overlap between the IPD and AgD studies. It means the resulting comparisons was conducted on a limited number of individuals in the IPD study and may be unstable.

Of the identified TAs, 4 did not attempt an indirect comparison [13, 17, 22, 24]. In TA529 [13], instead of using evidence from single-arm study, two RCTs were used to make relative treatment effect estimates as the company claimed that the treatments from the RCTs were similar to the intervention from the single-arm study.

In TA630 [22], no published data were available for making an indirect comparison as there was no comparator treatment available for the disease. Therefore, the company considers a population for comparison as the comparator arm that is in line with the existing standard of care.

In TA644 [24], the intervention study was a basket study that allowed investigators and manufacturers to conduct tumour-agnostic studies across multiple solid tumours by grouping cancer patients by their common genomic alterations. A formal indirect treatment comparison was not deemed feasible due to the significant heterogeneity between patient and disease characteristics, tumour types, and potential comparator therapies meeting the definition of 'established care'.

Table 2 summarises the application of different effect measures that was used in the TAs. Thirteen appraisals that have applied population adjustment with MAIC, more than half of them (9 out of 13, 69.23%) mentioned that the weighted Cox proportional hazard model has been used to estimate relative effectiveness [10, 15, 18, 19, 21, 23, 25, 27, 28]. Despite the fact that Cox proportional models do not specify the baseline hazard, still it is widely used.

Other than the weighted Cox model, the application of other measures has also been seen. In TA604 [20], a weighted survival function was used for making relative effectiveness. In TA716 [26], mean survival has been estimated using a semi-parametric model for treatment comparison as the PH assumption was very unlikely to hold for comparisons of the intervention and the comparator's treatment. Other than the weighted Cox proportional hazard model, weighted risk difference has also been used for treatment comparison [29].

Three TAs [11, 12, 14] have used STC with fractional polynomial network meta-analysis (FP NMA). In order to do the analysis, first, the company conducted a STC by incorporating bootstrapping to produce estimates of variability. A bootstrap sample is a random sample with replacement generated from the IPD in the intervention study. The company states that on average about 1/3 of the patients were not included in each bootstrap sample and called these patients out-of- bag (OOB). The Cox proportional hazards model was used to develop the regression model informed by baseline covariates. The company simulated a large number of hypothetical individuals based on the reported marginal distribution of the covariates of interest and the correlation from the intervention study. The company also generated the predicted log hazards.

The mean of the predicted log hazard and the variance of the log hazard from bootstrap samples were used in the fractional polynomial (FP) NMA model to produce time-varying hazard ratios for each comparator. The different kinds of effect measures using MAIC and STC have been summarised in Table 2.

All the appraisals of this review include unanchored comparisons without any common treatment, so the comparisons made are subject to unknown amounts of residual bias. Out of 16 TAs that have attempted unanchored comparison, only TA530 [14] has attempted to quantify residual bias by using out sample method which was one of the recommended methods from Phillippo et al. (2016).

3.3 Larger Network

Of these 20 appraisals, 11 of them (55%) have a larger network, i.e. either an intervention treatment has been compared to more than one comparator treatment where the comparator treatments have come from different studies or an intervention treatment has been compared to one comparator treatment but from multiple studies.

Table 2: Application of different effect measures

Effect measure for relative comparison	Number of appraisals
Weighted Cox proportional model (using MAIC weights)	9
Weighted survival function (using MAIC weights)	1
Weighted risk difference (using MAIC weights)	1
Weighted KM survival function (using MAIC weights)	1
Mean survival from parametric model (using MAIC weights)	1
Adjusted time varying HR from FP NMA (using STC)	3

Of these 11 studies that have larger networks [11, 12, 14, 10, 16, 13, 18, 23, 24, 25, 26,] six of them have used multiple MAICs. Despite of having a larger network, TA644 [24] have done unadjusted indirect comparison as it claimed that no adjustment was needed due to similarity between studies. The TA's that have used multiple MAIC to estimate relative treatment effect with multiple comparators, used the MAIC estimates as stand-alone estimates except one. TA571 [18] have done a standard pairwise meta-analysis using the MAIC estimates to estimate an overall, pooled estimate.

Three appraisals [11, 12, 14] have used STC to construct a predicted treatment arm for each single-arm study. This predicted arm forms a newly-connected network and then analyses were made with FP NMA. This approach gives the opportunity to make a coherent set of relative effect estimates, but, it adds another additional assumption that no difference exists in prognostic and effect modifiers among the single-arm studies included in the NMA.

The number of comparator treatments for these larger networks ranges from 1 to 6. Apart from the single-arm study of the intervention treatment, no connected network of evidence was found for the comparator treatments. The reasons for not getting a connected network were either that most of the comparator studies were also single-arm studies or that no common comparator was found from the RCTs.

3.4 Preference of other methods over the chosen method

Of the included TAs, three of them have discussed other methods apart from the method that have been applied. In TA522 [11], STC was applied with FP NMA. MAIC was discussed as an alternative to STC but it was not implemented. The rationale of not conducting MAIC was that the MAIC method requires access to relatively large and complete data sets; often this level of data granularity is not reported within published articles. The company states that one approach is not necessarily favoured over the other between STC and MAIC; STC with bootstrap has the benefit of allowing cross-validation and assessment of the model performance. In TA592 [19], both MAIC and STC were performed as scenario analysis for a single comparison, and for the base case, a naïve indirect comparison was applied. The company states that MAIC which has been driven by a small number of patients was a major factor in their decision to use the results of the STC as a key alternative scenario analysis within the cost-effectiveness modelling. However, given the uncertainties surrounding both of the indirect comparisons performed, the naive (unadjusted) comparison was used in the base case analysis as this provided the most conservative estimates of treatment effectiveness. In TA716 [26], MAIC was chosen over STC because though STC can be applied to a large number of comparators, for multiple outcomes, a outcome equation needs to be determined for each comparison which is often problematic specially for time-to-event data. As there was a single comparator but many outcomes to be compared, MAIC was a better option as after estimating the weights once, they could be used for multiple outcomes.

3.5 Uncertainty estimates in the population adjustment methods

The TAs included in this review either have conducted unanchored MAIC or STC for population adjustment. Most of the TAs did not mention what methods were used to take into account uncertainty. Five TAs that have used MAIC, mentioned sandwich estimator [10] and bootstrap [19, 21, 25, 27] were used to capture uncertainty. Three TAs that have used STC [11, 12, 19] mentioned bootstrap as their method to capture uncertainty.

3.6 Adjustment methods in survival extrapolation

For survival extrapolation, 12 TAs [10, 11, 12, 14, 16, 18, 21, 23, 25, 26, 27, 29] estimated the absolute survival effects directly based on the unadjusted Kaplan-Meier (KM) functions of the intervention study. The absolute effects for comparator treatments were estimated in the cost-effectiveness model by applying adjusted HR to the intervention KM function. For STC, time-varying hazard ratios (HR) which were estimated from the FP NMA model were applied to the KM function of the intervention study. In TA554 [16], an independent parametric model was fitted to intervention and comparator treatment without any adjustment. The company states that the 95% confidence intervals for both the adjusted and unadjusted intervention treatment curves were found to overlap for OS versus both comparators which indicates no difference in efficacy. As such, the company considered keeping the unadjusted OS profiles for intervention treatment in the base case economic analysis are more appropriate as it preserves patient numbers. In all these TAs, the target population of extrapolation was not clearly defined. Only one TA [20] applied a two-stage method, where, in the first stage, adjusted KM survival function was digitised and pseudo IPD were reconstructed for intervention study and in the second stage, parametric models were fitted separately for intervention and comparator study.

3.7 Identification and inclusion of covariates

All the adjustment methods in this review are in unanchored form, so all prognostic and effect modifier variables have to be included. For 10 out of the 20 TAs, the strategy for identification of variables was either literature search or a clinical expert opinion or a combination of

both [10, 11, 14,17, 18, 19, 25, 26, 28, 29]. The availability of final model was selected based on the best predictive performance. In TA530 [14] using STC, the identified variables were included in a Cox proportional hazard model by the stepwise model selection algorithm. In TA522 [11], model selection was based on the OOB predictive baseline characteristics reported in both studies was also one of the common approaches for variable selection [15, 20, 26]. Most of the appraisals did not discuss whether the identified variables were prognostic or effect modifiers. In TA643 [23]; TA554 [16]; TA529 [13]; TA722 [27] the company did not mention how they identified the important covariates. In TA510 [10], the company asked the clinical experts to rank covariates according to their relevance, then the covariates were added into the model in decreasing order of importance. The final model was determined by consideration of ESS.

In TA571 [18], the company gathered feedback from five clinicians through interviews and questionnaires where clinicians were asked to identify the variable which they believed to be important on survival outcomes. Clinicians were also asked to rank each variable according to its importance so that any trends across the clinician responses can be captured. In TA628 [21] and TA756 [29], variables were identified using both clinical feedback and cox regression. For these TAs, Cox regression was performed both in univariate and multivariate settings.

Three TAs that used STC for unanchored comparison, used statistical techniques to choose covariates. In TA525 [12], nine competing models fit were tested. The models were comprised of different combinations of covariates and their interaction terms and the final model was selected based on the best predictive performance. In TA530 [14] using STC, the identified variables were included in a Cox proportional hazard model by the stepwise model selection algorithm. In TA522 [11], model selection was based on the OOB predictive performance. The company defined the sum of Akaike Information Criterions (AICs) to be the sum of the differences between the observed KM survival estimates minus the predicted OOB survival estimates at every failure time in the original IPD KM curve. The model with the lowest sum of AICs would be chosen as the final model. If all models provide similar AICs, then the simplest model would be chosen.

Almost all the TAs except two [11, 12] did not include all the identified prognostic and effect modifier variables in the final model. The reason for not including all the variables was almost always a lack of availability issue. However, without including all prognostic and effect-modifying variables in the adjustment, the estimates will remain biased. In TA592 [19], variables were excluded from the final model due to convergence issues. When the algorithm did not converge that was used to estimate weights, variables were eliminated in a stepwise fashion in a pre-determined order until convergence is achieved. Out of 16 TAs which have used population adjustment methods, 10 have mentioned the number of variables they have adjusted for. The number of covariates adjusted for ranged from 2 to 14.

4. Discussion

It was found that population adjustment methods (MAIC and STC) were not always used as the company's base-case analysis when relative treatment effect estimated using single-arm study. They were included sometimes as supportive evidence. Some TAs conducted naive indirect comparisons despite important differences among the prognostic factors. In most of the TAs, the EAG expressed their concern about the fact that important baseline characteristics were not adjusted because of poor reporting by the studies. All identified prognostic and effect modifiers were not included in the final model due to their lack of availability in the included studies. This may cause residual confounding, which indicates that the populations being compared may still be considerably imbalanced and the impact of these imbalances on the survival estimates induces substantial uncertainty.

When STC is applied for time-to-event outcomes, the treatment effect is non-collapsible which means the marginal and conditional effects do not coincide even if there is no confounding and the distribution of covariates is balanced. As a result, STC will produce a systematic bias for time-to-event outcomes [8]. TA525 [12] that have used STC and FP NMA to deal with a larger network of evidence, a further issue was identified which relate to the choice of comparator studies. The level of heterogeneity was found to be moderate to high in this appraisal.

Although MAIC was found to be the most applied population adjustment method, a considerable re- duction was seen in ESS. Adjusting for all baseline imbalances was often discarded in order to avoid a substantial loss in sample size. The issue of small ESS depends on the original sample size. If the original sample size is big, then a reduction ESS may still be enough to conduct a MAIC. TA604 [20] had a considerable reduction of the sample size which resulted in an ESS of 3.8,i.e. 94.73% reduction in sample size. Very small ESS indicates that there was a serious lack of overlap between the IPD and AgD populations which makes the MAIC results unreliable. When there is a lack of overlap, weighting methods like MAIC are unable to extrapolate beyond those observed in the IPD, and may produce an estimate which remains biased. The uncertainty estimates of the MAIC methods also remain unclear as most of the TAs did not provide information on this. All the MAICs found

in this review did not mention whether the propensity score model includes second-order terms or not. Several TAs found in this review deal with multiple comparator studies with AgD, which means multiple comparisons were required. Current MAIC and STC originally target a single comparison and if a larger network needs to be dealt with, an additional assumption called "shared effect modifier" needs to be satisfied. TAs that have used MAIC for larger networks have conducted multiple MAICs. Each of these comparisons is valid for different target populations and none of the TAs has tried to justify a coherent analysis with a shared effect modifier assumption. Moreover, during the multiple MAICs, IPD from the intervention study was used multiple times which will create additional complexity as independence between the studies was broken. It creates a correlation between the estimates of treatment effects which remain unaddressed. TAs that used STC and then NMA, have tried to make a coherent synthesis which requires further assumptions in the process, that is, prognostic, as well as effect modifiers, are balanced across the studies. This assumption was questionable when moderate to high heterogeneity was found [12]. The issues found with a considerable reduction of MAIC ESS and dealing with a larger network of evidence are similar to the findings by Phillippo et al. (2019b). Although Phillippo et al. (2016) have identified that current MAIC and STC is unable to be implemented for larger network with- out making an additional assumption, some TAs are still using them for a larger network.

All the MAIC and STC included in this review are in an unanchored form which assumes that there are no unmeasured prognostic and effect modifier factors. This assumption is very hard to justify and therefore some suggestion by Phillippo et al. (2016) was given for quantifying residual bias due to unmeasured confounding. Except for TA 530, no other TAs have tried to follow these suggestions. TA 530 has used out-sample method to quantify the residual bias. A question was included in the data extraction sheet to highlight whether included studies had "immature data" or not, to assess the appropriateness of extrapolation and the ability to adjust for covariates. No data was available from the included TAs but in most of the TAs, EAG has expressed concern about data immaturity.

Extrapolation with time-to-event data can be done in several ways. First, after adjusting time-to-event outcomes, adjusted parametric models can be fitted separately for intervention and comparator study or a joint model can be fitted including treatment as a covariate. Second, the adjusted KM survival function can be digitised and pseudo IPD can be reconstructed for intervention study and then parametric models can be fitted for intervention and comparator study. In these approaches, the population of extrapolation is the AgD study. Another strategy of extrapolation is that the parametric model is fitted to the unadjusted IPD of the intervention study. The absolute effects for comparator treatments are then estimated in the cost- effectiveness model by applying adjusted HR to the intervention KM function. In this case, the population of extrapolation is the IPD study.

The most common strategy of extrapolation was to fit a parametric model with the unadjusted IPD of the intervention study. The fitted model will give survival probabilities of the intervention treatment in the population of the intervention study. The survival extrapolation for comparator studies were estimated in the cost-effectiveness model by applying MAIC/STC adjusted HR to these survival probabilities. It provides the survival probabilities for the comparator treatment in the population of the intervention study. When extrapolation is done directly to the unadjusted KM function of the intervention study and comparator treatment extrapolation is obtained by applying HR, the population of extrapolation is the IPD population. This can only be done if treatment effects are transportable.

None of the TAs included in this review has given any justification for the transportable treatment effect. Additionally, if proportional hazard (PH) assumption was made when applying HR, it should be tested for both unadjusted and adjusted comparisons. None of the TAs have tested this assumption in both case. Only two TAs [10,25] have tested for unadjusted and three TAs [18, 27, 28] have tested for adjusted comparison. Using this extrapolation approach, a PH/AFT assumption is made which assumes a constant treatment effect. It is very important to consider whether this is likely to hold over the entire time be in oncology where survival outcomes (e.g. progression-free survival, overall survival) were found to be most common outcome type. MAIC and STC found to be mostly used methods which is true for this review also. A horizon. None of the TAs have made any attempt to justify this assumption. As this is unlikely to be true, so this approach is not appreciated generally.

TA604 [20] have applied a two-stage extrapolation approach where MAIC adjusted KM survival function for intervention study is digitised and pseudo patient level data are reconstructed and in the second stage, parametric models are fitted separately for intervention and comparator study. A check was con- ducted to ensure good overlapping between the weighted KM survival function and the reconstructed KM survival function to ensure that the algorithm by Guyot et al. (2012) has reasonably worked well. How-ever, as the overlapping

was found to be very poor, a sample inflation method was implemented where analyst judgment was used to increase the number-at-risk at time zero iteratively until the recreated KM curves provided a visually good fit to the original KM curves. As a consequence of this sample-inflation approach, the variance and hence confidence intervals around parameters of parametric survival models fitted to these data were underestimated. All the issues identified with survival extrapolation were found in addition to Phillippo et al. (2019b) review.

Many of the findings of this review matches with that of Phillipo's review. Like this review, most of the applications of population adjustment in Phillipo's review found to be substantial decrease in ESS has been found in both of the reviews which in turn made the comparisons dependent on very few number of individuals in the IPD study. In Phillipo's review no TAs found to be attempted to adjust residual bias but in this review one TA found to be attempted out of sample method. The range for covariate adjustment is also found to be similar. The number of covariates adjusted for are 2 to 14 and 1 to 13 for this and Phillipo's review respectively.

Review by Phillippo et al. (2019b) also found that the most common approach to identifying prognostic and effect modifier variables was their availability in the included studies. In this review, although the same approach was implemented by some TAs, the majority of the TAs have used literature review and clinical expert opinion to identify variables. This shows that TAs are trying to follow guidance on variable selection by Phillippo et al. (2016). Furthermore, the application of population adjustment methods were found to be very prevalent for larger network of evidence(10 out of 18 TAs (55.6%)) which is also true for this review.

5. Conclusion

It is not surprising that adjustment methods like unanchored MAIC and STC were frequently used in NICE TAs with the increase of single-arm studies. These methods assume no unmeasured effect modifiers and prognostic factors which is difficult to satisfy and in turn can produce residual bias. In addition, these methods target the comparator population which is often not the population of interest. Moreover, these methods were frequently used for larger networks despite the fact that they are not developed to handle multiple comparator treatments simultaneously.

6. Conflict of interest

The author affirms that there is no conflict of interest regarding the publication of this paper.

References

1. Cucherat, M., Laporte, S., Delaitre, O., Behier, J.-M., d'Andon, A., Binlich, F., Bureau, S., Cornu, C., Fouret, C., Labouret, N. H., et al. (2020). From single-arm studies to externally controlled studies: Methodological considerations and guidelines. Therapies, 75(1), 21-27.
2. Dias, S., Sutton, A. J., Ades, A., & Welton, N. J. (2013). Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical Decision Making, 33(5), 607-617.
3. Guyot, P., Ades, A., Ouwens, M. J., & Welton, N. J. (2012). Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology, 12(1), 9.
4. National Institute for Health and Care Excellence (NICE). (2009). Guide to the single technology appraisal process. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/Guide-to-the-single-technology-appraisal-process.pdf
5. Phillippo, D., Ades, T., Dias, S., Palmer, S., Abrams, K. R., & Welton, N. (2016). NICE DSU technical support document 18: Methods for population-adjusted indirect comparisons in submissions to NICE.
6. Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K. R., & Welton, N. J. (2018). Methods for population-adjusted indirect comparisons in health technology appraisal. Medical Decision Making, 38(2), 200-211.
7. Phillippo, D. M., Dias, S., Elsada, A., Ades, A., & Welton, N. J. (2019a). Population adjustment methods for indirect comparisons: A review of National Institute for Health and Care Excellence technology appraisals. International Journal of Technology Assessment in Health Care, 35(3), 221-228.
8. Phillippo, D. M., Dias, S., Welton, N. J., Caldwell, D. M., Taske, N., & Ades, A. (2019b). Threshold analysis as an alternative to GRADE for assessing confidence in guideline recommendations based on network meta-analyses. Annals of Internal Medicine, 170(8), 538-546.
9. Remiro-Azócar, A., Heath, A., & Baio, G. (2020). Methods for population adjustment with limited access to individual patient data: A simulation study. arXiv preprint arXiv:2004.14800.
10. Sealey, S., Edathodu, A., & Mukku, S. R. (2014). International HTA referencing-a reality? Value in Health, 17(3), A28.

11. National Institute for Health and Care Excellence (NICE). (2018). Daratumumab monotherapy for treating relapsed and refractory multiple myeloma (TA510). https://www.nice.org.uk/guidance/ta783/documents/committee-papers
12. National Institute for Health and Care Excellence (NICE). (2018). Pembrolizumab for locally advanced or metastatic urothelial cancer is cisplatin where unsuitable (TA522). https://www.nice.org.uk/guidance/ta522/documents/committee-papers
13. National Institute for Health and Care Excellence (NICE). (2018). Atezolizumab for treating metastatic urothelial cancer after platinum-based chemotherapy (TA525). https://www.nice.org.uk/guidance/ta525/documents/committee-papers-2
14. National Institute for Health and Care Excellence (NICE). (2018). Crizotinib for treating ROS1-positive advanced non-small-cell lung cancer (TA529). https://www.nice.org.uk/guidance/ta529/documents/committee-papers-2
15. National Institute for Health and Care Excellence (NICE). (2018). Nivolumab for treating metastatic or unresectable urothelial cancer after platinum-based chemotherapy (TA530). https://www.nice.org.uk/guidance/ta530/documents/committee-papers-3
16. National Institute for Health and Care Excellence (NICE). (2018). Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma (TA540). https://www.nice.org.uk/guidance/ta540/documents/committee-papers-2
17. National Institute for Health and Care Excellence (NICE). (2018). Tisagenlecleucel for treating relapsed or refractory B-cell acute lymphoblastic leukaemia in people aged up to 25 years (TA554). https://www.nice.org.uk/guidance/ta554/documents/committee-papers
18. National Institute for Health and Care Excellence (NICE). (2019). Tisagenlecleucel-T for treating relapsed or refractory diffuse large B-cell lymphoma(TA567). https://www.nice.org.uk/guidance/ta567/documents/committee-papers
19. National Institute for Health and Care Excellence (NICE). (2019). Brigatinib for treating ALK-positive non-small-cell lung cancer after crizotinib (TA571). https://www.nice.org.uk/guidance/ta571/documents/committee-papers
20. National Institute for Health and Care Excellence (NICE). (2019). Cemiplimab for treating metastatic or locally advanced cutaneous squamous cell carcinoma (TA592). https://www.nice.org.uk/guidance/ta802/documents/committee-papers
21. National Institute for Health and Care Excellence (NICE). (2019). Idelalisib for treating follicular lymphoma refractory to 2 treatments (TA604). https://www.nice.org.uk/guidance/ta604/documents/committee-papers
22. National Institute for Health and Care Excellence (NICE). (2020). Lorlatinib for previously treated ALK-positive advanced non-small-cell lung cancer (TA628). https://www.nice.org.uk/guidance/ta628/documents/committee-papers-2
23. National Institute for Health and Care Excellence (NICE). (2020). Larotrectinib for treating NTRK fusion-positive advanced solid tumours (TA630). https://www.nice.org.uk/guidance/ta630/documents/committee-papers
24. National Institute for Health and Care Excellence (NICE). (2020). Entrectinib for treating ROS1 fusion-positive locally advanced or metastatic non-small-cell lung cancer (TA643). https://www.nice.org.uk/guidance/ta643/documents/committee-papers
25. National Institute for Health and Care Excellence (NICE). (2020). Entrectinib for treating NTRK fusion-positive solid tumours (TA644). https://www.nice.org.uk/guidance/ta644/documents/committee-papers
26. National Institute for Health and Care Excellence (NICE). (2021). Trastuzumab deruxtecan for treating HER2-positive unresectable or metastatic breast cancer after 2 or more anti-HER2 therapies (TA704). https://www.nice.org.uk/guidance/ta704/documents/committee-papers
27. National Institute for Health and Care Excellence (NICE). (2021). Nivolumab with ipilimumab for previously treated metastatic colorectal cancer with high microsatellite instability or mismatch repair deficiency (TA716). https://www.nice.org.uk/guidance/ta716/documents/committee-papers
28. National Institute for Health and Care Excellence (NICE). (2021). Pemigatinib for treating relapsed or refractory advanced cholangiocarcinoma with FGFR2 alterations (TA722). https://www.nice.org.uk/guidance/ta722/documents/committee-papers-3
29. National Institute for Health and Care Excellence (NICE). (2021). Selpercatinib for treating advanced thyroid cancer with RET alterations (TA742). https://www.nice.org.uk/guidance/ta742/documents/committee-papers

30. National Institute for Health and Care Excellence (NICE). (2021). Fedratinib for disease-related splenomegaly and symptoms in myelofibrosis (TA756). https://www.nice.org.uk/guidance/ta756/documents/committee-papers

31. National Institute for Health and Care Excellence (NICE). (2022). Pralsetinib for treating RET fusion-positive advanced non-small-cell lung cancer (TA812). https://www.nice.org.uk/guidance/ta812/documents/committee-papers-2

32. National Institute for Health and Care Excellence (NICE). (2023). Zanubrutinib for treating chronic lymphocytic leukaemia (TA931). https://www.nice.org.uk/guidance/ta931/documents/committee-papers-2

Appendix Table 1

1.	Name/TA number
2.	Publication date of the appraisal.
3.	Protocol number and/or name of the pivotal study.
4.	Therapeutic area.
5.	What types of outcome measures were indirectly compared?
6.	What type of network is being considered?
7.	What methodology was used to conduct unanchored indirect comparison?
8.	How were covariates included in the model?
9.	How many variables were included in the model?
10.	Were all identified prognostic and effect modifier variables included in the model?
11.	If not, what was the reason for excluding variables?
12.	Other than prognostic and effect modifier variables, were other variables also included?
13.	Were second-order terms included?
14.	What were the original sample sizes?
15.	What was the subsequent sample size?
16.	If NMA was conducted, was any attempt made to check if any inconsistency found in the connected part of the network?
17.	Was heterogeneity among the studies assessed?
18.	If yes, what was the amount of heterogeneity identified?
19.	Were at least two studies available on the on each contrast for the heterogeneity parameter?
20.	Along with the chosen model, were other method also discussed?
21.	Was any justification given for the chosen model?
22.	How many events were available for the time-to-event outcomes?
23.	What approach was used for the extrapolation of time-to-event data?
24.	What adjustment was made for time-to-event data?
25.	Is overlapping between weighted and reconstructed KM been checked/commented on?

26.	If not, what procedure was taken to ensure overlapping between weighted IPD and reconstructed IPD?
27.	Was the population for the extrapolation clearly defined?
28.	Treatment effect was estimated for which population?
29.	Had any justification given for transportable treatment effect if they were estimated for IPD study?
30.	If PH assumption was made, was it tested both for the adjusted and unadjusted population?
31.	What procedure has been taken to measure uncertainty?
32.	What attempt was made to quantify residual bias?

Appendix Table A.1:: NICE STA data extraction table

TA number	TA 525	TA 604	TA 510
Name of the pivotal study	IMvigor210	study 101-09	MMY2002
Publication date of the appraisal	13/06/2018	02/10/2019	14/03/2018
Therapeutic Area	oncology	oncology	oncology
Outcome measures indirectly compared	Time-to-event	Time-to-event	Time-to-event
Which timetoevent compared	OS	OS+PFS	OS+PFS
Type of network	Multiple comparison (large)	Single comparison	Multiple comparison (large)
Methodology for unanchored comparison	NMA by STC	MAIC	MAIC
How covariates identified	predictive performance	availability of both study	literature review + clinical expert
How many variables included	4	not mentioned	MAIC1=11, MAIC2=5
Were all prognostic/effect modifiers included?	yes	no	no
If no, reason	not mentioned	lack of availability	lack of availability

Other variables included?	no	no	no
Only main effects?	yes	not mentioned	not mentioned
Secondorder terms?	no	not mentioned	not mentioned
MAIC effective sample size (%)	not applicable	3.8 (5.27%)	MAIC1=84(56.75%), MAIC2=80(64%)
NMA inconsistency checked?	no	not applicable	not applicable
Heterogeneity assessed?	yes	not applicable	not applicable
Amount of heterogeneity	moderate	not applicable	not applicable
Two studies available for contrasts?	not mentioned	not applicable	not applicable
Other methods discussed?	yes	no	no
Justification for chosen method?	yes	not applicable	not applicable
Events available for TTE outcome	not applicable	not available	not available
CEA extrapolation approach	parametric model (unadjusted survival)	two stage	parametric model (unadjusted survival)
Clinical effectiveness adjustment	adjusted time-varying HR from NMA	weighted KM survival	weighted Cox PH model
Overlap between weighted IPD & reconstructed IPD checked?	not applicable	yes	not applicable
If no, procedure	not applicable	sample inflation approach	not applicable
Population for extrapolation defined?	no	yes	no
Population for treatment effect	IPD	AgD	IPD
Justification for transportability?	no	not applicable	no
PH assumption tested?	not applicable	not applicable	for unadjusted
Procedure to measure uncertainty	bootstrap	not mentioned	sandwich estimator
Attempt to estimate residual bias?	no	no	no

Appendix Table A.2:: NICE STA data extraction table

TA number	TA 530	TA 628	TA 643
Name of the pivotal study	CheckMate 275 and CheckMate 032	Study 1001	ALKA, STARTRK-1, STARTRK-2
Publication date of appraisal	04/07/2018	13/05/2020	12/08/2020
Therapeutic Area	oncology	oncology	oncology
Outcome measures indirectly compared	Time-to-event	Time-to-event	Time-to-event
Which time-to-event	OS+PFS	OS+PFS	OS+PFS
Type of network	Multiple comparison (large)	Single comparison	Multiple comparison (large)
Methodology for unanchored comparison	NMA by STC	MAIC	MAIC
How covariates identified	literature review + clinical expert	clinical feedback + Cox regression (uni + multi)	not mentioned
How many variables included	4 out of 11	4	MAIC1=6, MAIC2=6
All prognostic/effect modifiers included?	no	no	no
If no, reason	lack of availability	lack of availability	lack of availability
Other variables included?	no	no	no
Only main effects?	not mentioned	not mentioned	not mentioned
Second-order terms?	not mentioned	not mentioned	not mentioned
MAIC effective sample size (%)	not applicable	not mentioned	not mentioned
NMA inconsistency checked?	no	not applicable	not applicable
Heterogeneity assessed?	no	not applicable	not applicable
Amount of heterogeneity	not mentioned	not applicable	not applicable
Two studies per contrast?	no	not applicable	not applicable
Other methods discussed?	no	yes	no
Justification for chosen method?	no	no	no
Events available for TTE	not available	not available	not available

CEA extrapolation method	parametric model (unadjusted survival)	parametric model (unadjusted survival)	parametric model (unadjusted survival)
Clinical effectiveness adjustment	adjusted time-varying HR (NMA)	weighted Cox PH model	weighted Cox PH model
Overlap between weighted & reconstructed IPD checked?	not applicable	no	no
If no, overlapping procedure	not applicable	no	no
Population for extrapolation defined?	no	no	no
Population for treatment effects	IPD	IPD	IPD
Justification for transportability?	no	no	no
PH assumption tested?	not applicable	not mentioned	not mentioned
Uncertainty measurement	not mentioned	bootstrap	not mentioned
Residual bias estimated?	out sample	no	not mentioned

Appendix Table A.3:: NICE STA data extraction table

TA number	TA 554	TA 567	TA 571
Name of the pivotal study	ENSIGN, ELIANA and B2101J	JULIET trial	ALTA and Study 101
Publication date of the appraisal	21/12/2018	13/03/2019	20/03/2019
Therapeutic Area	Oncology	Oncology	Oncology fv
Outcome measures compared	Time-to-event	Time-to-event	Time-to-event
Which TTE compared	EFS+OS	OS+PFS	OS+PFS
Type of network	Multiple comparison	Single comparison	Multiple comparison
Methodology	MAIC	Unadjusted indirect comparison	MAIC
How covariates identified	not mentioned	expert opinion	expert opinion

Variables included	not mentioned	8	not mentioned
All prognostic/effect modifier included?	no	no	no
Reason if no	lack of availability	lack of availability + variation among clinicians	lack of availability
Other variables included?	not mentioned	not mentioned	no
Only main effects?	not mentioned	not applicable	not mentioned
Second order terms?	not mentioned	not applicable	not mentioned
MAIC ESS (%)	not mentioned	not applicable	MAIC1=67.1 (49.70%), MAIC2=76.5 (56.66%)
NMA inconsistency checked?	not applicable	not applicable	not applicable
Heterogeneity assessed?	not applicable	not applicable	not applicable
Amount of heterogeneity	not applicable	not applicable	not applicable
2 studies per contrast?	not applicable	not applicable	not applicable
Other methods discussed?	no	no	no
Justification for method?	not applicable	not applicable	not applicable
Events available (%)	not available	not available	not available
Approach for extrapolation	Independent parametric model (unadjusted)	Independent parametric model (unadjusted)	parametric model (unadjusted)
Adjustment in clinical effectiveness	weighted Cox PH model	not applicable	weighted Cox PH model
Overlap checked?	not applicable	not applicable	not applicable
Procedure to ensure overlap	not applicable	not applicable	not applicable
Population for extrapolation	no	not applicable	no
Population for treatment effects	IPD	not mentioned	IPD
Justification for transportability	not mentioned	no	not mentioned

PH tested?	not mentioned	not applicable	For adjusted comparisons only
Uncertainty measurement	not mentioned	not applicable	not mentioned
Residual bias assessed?	not mentioned	not applicable	not mentioned

Appendix Table **A.3:**: *NICE STA data extraction table*

TA number	TA 554	TA 567	TA 571
Name of the pivotal study	ENSIGN, ELIANA and B2101J	JULIET trial	ALTA and Study 101
Publication date of the appraisal	21/12/2018	13/03/2019	20/03/2019
Therapeutic Area	Oncology	Oncology	Oncology fv
Outcome measures compared	Time-to-event	Time-to-event	Time-to-event
Which TTE compared	EFS+OS	OS+PFS	OS+PFS
Type of network	Multiple comparison	Single comparison	Multiple comparison
Methodology	MAIC	Unadjusted indirect comparison	MAIC
How covariates identified	not mentioned	expert opinion	expert opinion
Variables included	not mentioned	8	not mentioned
All prognostic/effect modifier included?	no	no	no
Reason if no	lack of availability	lack of availability + variation among clinicians	lack of availability
Other variables included?	not mentioned	not mentioned	no
Only main effects?	not mentioned	not applicable	not mentioned
Second order terms?	not mentioned	not applicable	not mentioned
MAIC ESS (%)	not mentioned	not applicable	MAIC1=67.1 (49.70%), MAIC2=76.5 (56.66%)
NMA inconsistency checked?	not applicable	not applicable	not applicable
Heterogeneity assessed?	not applicable	not applicable	not applicable
Amount of heterogeneity	not applicable	not applicable	not applicable
2 studies per contrast?	not applicable	not applicable	not applicable

Other methods discussed?	no	no	no
Justification for method?	not applicable	not applicable	not applicable
Events available (%)	not available	not available	not available
Approach for extrapolation	Independent parametric model (unadjusted)	Independent parametric model (unadjusted)	parametric model (unadjusted)
Adjustment in clinical effectiveness	weighted Cox PH model	not applicable	weighted Cox PH model
Overlap checked?	not applicable	not applicable	not applicable
Procedure to ensure overlap	not applicable	not applicable	not applicable
Population for extrapolation	no	not applicable	no
Population for treatment effects	IPD	not mentioned	IPD
Justification for transportability	not mentioned	no	not mentioned
PH tested?	not mentioned	not applicable	For adjusted comparisons only
Uncertainty measurement	not mentioned	not applicable	not mentioned
Residual bias assessed?	not mentioned	not applicable	not mentioned

Appendix Table A.4:: NICE STA data extraction table

TA number	TA 522	TA 529	TA 540
Name of the pivotal study	KEYNOTE-052	PROFILE 1001, PROFILE 1014, PROFILE 1007	KEYNOTE-087
Publication date of the appraisal	13/06/2018	04/07/2018	03/09/2018
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event
Which time-to-event was indirectly compared	EFS+OS	OS+PFS	PFS
What type of network is being considered?	Multiple comparison (larger network)	Multiple comparison (larger network)	Single comparison

Methodology used for unanchored indirect comparison	NMA by STC	No indirect comparison	MAIC
How covariates were identified	model predictive performance + literature review + expert opinion	not applicable	availability
How many variables included in model	5	not applicable	not mentioned
All prognostic/effect modifier included?	yes	not applicable	no
If no, reason	not applicable	not applicable	lack of availability
Other variables included?	no	not applicable	not mentioned
Main effects only?	yes	not applicable	not mentioned
Second order terms?	no	not applicable	not mentioned
MAIC effective sample size (%)	not applicable	not applicable	not mentioned
NMA inconsistency checked?	not mentioned	not applicable	not applicable
Heterogeneity assessed?	not mentioned	not applicable	not applicable
Residual bias estimated?	not mentioned	not applicable	not mentioned

Appendix Table A.5:: NICE STA data extraction table

TA number	TA630	TA644	TA592
Name of the pivotal study	LOXO-TRK-14001, NAVIGATE (LOXO-TRK-15002) and SCOUT (LOXO-TRK-15003)	(ALKA, STARTRK-1, STARTRK-2, and STARTRK-NG)	EMPOWER-CSCC 1
Publication date of the appraisal	27/05/2020	12/08/2020	07/08/2019
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event
Which time-to-event was indirectly compared	EFS + OS	OS + PFS	PFS
What type of network is being considered?	Multiple comparison (larger network)	Multiple comparison (larger network)	Single comparison
What methodology used for making unanchored indirect comparison	unadjusted comparison	unadjusted comparison	MAIC + STC
How covariates were identified	not applicable	not applicable	literature review + clinical expert

How many variables were included in the model	not applicable	not applicable	2 out of 12
Were all identified prognostic and effect modifier variables included in the model	not applicable	not applicable	no
If no, what was the reason	not applicable	not applicable	not mentioned
Other than prognostic and effect modifier variables, were other variables also included	not applicable	not applicable	no
Did the model only include main effects	not applicable	not applicable	not mentioned
Did the model include second order terms	not applicable	not applicable	not mentioned
MAIC effective sample size (%)	not applicable	not applicable	37 (34.3%)
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	not applicable	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies assessed?	not applicable	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies assessed?	not applicable	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	not applicable	not applicable	not applicable
Was at least two studies available on each contrast for the heterogeneity parameter	not applicable	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	no	no	yes
Was any justification given for the chosen method?	no	no	yes
How many events were available for time-to-event outcome (%)	not available	not available	not available
In cost effectiveness, what approach was used for extrapolation of time-to-event data	parametric model with unadjusted survival function	parametric model with unadjusted survival function	one stage
In clinical effectiveness, what adjustment was made for time-to-event data	adjusted time-varying HR from NMA	not applicable	weighted Cox proportional hazard model
Is overlapping between weighted IPD and reconstructed IPD checked/commented on	not applicable	not applicable	not applicable

If no, what procedure was taken to ensure overlapping between weighted IPD and reconstructed IPD	not applicable	not applicable	not applicable
Was the population for the extrapolation clearly defined?	not applicable	not applicable	yes
Treatment effects are estimated for which population?	not mentioned	not mentioned	AgD
Had any justification given for transportable treatment effects if they are estimated for IPD population?	not applicable	not applicable	not mentioned
If PH assumption was made, was it tested?	not mentioned	not mentioned	not mentioned
What procedure has been taken to measure uncertainty	not mentioned	not mentioned	bootstrap
Any attempt made to estimate residual bias?	not mentioned	not mentioned	no

Appendix Table A.6:: NICE STA data extraction table

TA number	TA 704	TA 716	TA 722
Name of the pivotal study	DESTINY-Breast01	CheckMate 142	FIGHT-202
Publication date of the appraisal	27/02/2021	28/07/2021	25/08/2021
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event
Which time-to-event was indirectly compared	PFS + OS	OS + PFS	PFS
What type of network is being considered?	Multiple comparison (larger network)	Multiple comparison (larger network)	Single comparison
What methodology used for making unanchored indirect comparison	MAIC	MAIC	MAIC
How covariates were identified	literature review + clinical expert	availability + clinical expert	no justification
How many variables were included in the model	8	14	4
Were all identified prognostic and effect modifier variables included in the model	no	not applicable	no
If no, what was the reason	lack of availability	lack of availability	not mentioned
Other than prognostic and effect modifier variables, were other variables also included	no	no	no
Did the model only include main effects	not mentioned	not mentioned	not mentioned
Did the model include second order terms	not mentioned	not mentioned	not mentioned

MAIC effective sample size (%)	not available	not available	not available
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	not applicable	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies assessed?	not applicable	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	not applicable	not applicable	not applicable
Was at least two studies available on each contrast for the heterogeneity parameter	not applicable	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	no	yes	no
Was any justification given for the chosen method?	not applicable	yes	no
How many events were available for time-to-event outcome (%)	not available	not available	not available
In cost effectiveness, what approach was used for extrapolation of time-to-event data	Independent parametric model with unadjusted survival function	Independent parametric model with unadjusted survival function	Independent parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time-to-event data	weighted Cox proportional hazard model	mean survival from parametric model	weighted Cox proportional hazard model
Is overlapping between weighted IPD and reconstructed IPD checked/commented on	not mentioned	not applicable	not applicable
If no, what procedure was taken to ensure overlapping between weighted IPD and reconstructed IPD	not mentioned	not mentioned	not applicable
Was the population for the extrapolation clearly defined?	no	yes	no
Treatment effects are estimated for which population?	IPD	IPD	IPD
Had any justification given for transportable treatment effects if they are estimated for IPD population?	not mentioned	no	no
If PH assumption was made, was it tested?	for unadjusted comparison	not applicable	for adjusted comparison
What procedure has been taken to measure uncertainty	bootstrap	not mentioned	bootstrap
Any attempt made to estimate residual bias?	not mentioned	yes	not mentioned

Appendix Table **A.7:**: *NICE STA data extraction table*

TA number	TA 742	TA 756
Name of the pivotal study	LIBRETTO-001	142
Publication date of the appraisal	03/11/2021	16/12/2021
Therapeutic Area	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Binary
Which time-to-event was indirectly compared	PFS+OS	SVR+TSS
What type of network is being considered?	single comparison	single comparison
What methodology used for making unanchored indirect comparison	MAIC	MAIC+STC
How covariates were identified	literature review + clinical expert	clinical expert and univariable + multivariable analysis
How many variables were included in the model	6	3
Were all identified prognostic and effect modifier variables included in the model	no	no
If no what was the reason	lack of availability	lack of availability
Other than prognostic and effect modifier variables, were other variables also included	not mentioned	not mentioned
Did the model only include main effects	not mentioned	not mentioned
Did the model include second order terms	not mentioned	not mentioned
MAIC effective sample size(%)	not available	34.4(35.5%)
If NMA was conducted, any attempt made to check inconsistencies	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies assessed?	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	not applicable	not applicable
Was at least two studies available on each contrast for heterogeneity parameter	not applicable	not applicable

Along with the chosen method, were other methods also discussed?	no	yes
Was any justification given for the chosen method?	no	yes
How many events were available for time to event outcome(%)	not available	not available
In cost effectiveness, what approach was used for extrapolation	one stage	parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time to event data	weighted Cox proportional hazard model	weighted risk difference
Is overlapping between weighed IPD and reconstructed IPD checked/commented	not mentioned	not applicable
If no, what procedure was taken to ensure overlapping	not applicable	not applicable
Was the population for extrapolation clearly defined?	no	no
Treatment effects are estimated for which population?	AgD	IPD
Had any justification given for transportable treatment effects if estimated for IPD population?	not mentioned	not mentioned
If PH assumption was made, was it tested?	for adjusted comparison	no
What procedure has been taken to measure uncertainty	not mentioned	not mentioned
Any attempt made to estimate residual bias?	not mentioned	yes