s 22

From: s 47F
Sent: Tuesday, 19 November 2019 8:26 AM
To: s 22
Subject: Re: Are u around this week [SEC=OFFICIAL]

Hi s 22
That's fine and agree with red.
s 47F

from iphone
Pardon typos and shorthand

From: s 22
Sent: Thursday, November 14, 2019 6:10 pm
To: s 47F
Subject: RE: Are u around this week [SEC=OFFICIAL]

s 47F

Apologies for the delay.

Please find attached a "clean" version of the PSD s 47C, s 47E

s 47C, s 47E

Happy to discuss if needed when most convenient to you.

s 22
Office of HTA/Technology Assessment and Access Division

Department of Health
s 22
GPO Box 9848, Canberra ACT 2601
s 22

From: s 47F
Sent: Wednesday, 13 November 2019 4:09 PM
To: s 22
Subject: Re: Are u around this week [SEC=OFFICIAL]

s 47F from iphone
Pardon typos and shorthand

From: s 22
Sent: Wednesday, November 13, 2019 4:00 pm
To: s 47F
Subject: RE: Are u around this week [SEC=OFFICIAL]

Yes

Update on Oncotype DX:

- I have got s 47F
- s 47E s 47F s 47E
- s 47E

s 22

s 22

s 22

Office of HTA/Technology Assessment and Access Division
Department of Health
s 22
GPO Box 9848, Canberra ACT 2601
s 22

From: s 47F
Sent: Wednesday, 13 November 2019 3:56 PM
To: s 22
Subject: Are u around this week [SEC=No Protective Marking]

s 47F

Apologies for typos ipad message

"Important: This transmission is intended only for the use of the addressee and may contain confidential or legally privileged information. If you are not the intended recipient, you are notified that any use or dissemination of this communication is strictly prohibited. If you receive this transmission in error please notify the author immediately and delete all copies of this transmission."

Australian Government

Medical Services Advisory Committee

Public Summary Document

Application No. 1342.5 Gene expression profiling of 21 genes in breast cancer to quantify the risk of disease recurrence and predict adjuvant chemotherapy benefit

Applicant: **Specialised Therapeutics Australia Pty Ltd**

Date of MSAC consideration: **MSAC 76th Meeting, 1-2 August 2019**

Context for decision: MSAC makes its advice in accordance with its Terms of Reference, visit the MSAC website

1. Purpose of application

A resubmission seeking public funding for the gene expression profiling (GEP) test using the real-time reverse-transcriptase polymerase chain reaction (RT-PCR) technique for 21 genes (Oncotype DX® or ODX) in women with newly diagnosed stage I or II breast cancer, who are oestrogen receptor positive (ER-positive) or progesterone receptor positive (PR-positive), Human Epidermal Growth Factor Receptor 2 negative (HER2-negative), and lymph node negative (LN-negative), was received from Specialised Therapeutics by the Department of Health.

2. MSAC's advice to the Minister - August 2019

After considering the strength of the available evidence in relation to comparative safety, clinical effectiveness and cost-effectiveness, MSAC did not support public funding for this gene expression profiling test for patients with breast cancer primarily because its ability to identify those who could safely be spared the addition of chemotherapy to endocrine therapy was not demonstrated by the new trial. The re-analysis of previously provided evidence was also insufficient to change the previous conclusion that the test could not satisfactorily identify those intermediate–risk patients who would benefit from the addition of chemotherapy to endocrine therapy.

3. Summary of consideration and rationale for MSAC's advice

s 47C, s 47E

s 47C, s 47E

s 47C, s 47E

s 47C, s 47E

s 47C, s 47E

s 47C, s 47E

s 47C

4. Background

The original application (Application 1345) was considered by MSAC at its July 2013 meeting, subsequent resubmissions were then considered in April 2014, November 2015, July 2016 and July 2017. The PSDs for these applications can be viewed on the MSAC website.

At its July 2017 meeting, MSAC did not support Oncotype DX breast cancer assay due to the uncertainty of the incremental benefit of the Oncotype DX breast cancer assay over optimal care (Application 1342.4 Public Summary Document (PSD) 2017, p2). MSAC noted that data from ongoing trials like the TAILORx trial, if suitable, may be useful in addressing this uncertainty (PSD, p3).

5. Prerequisites to implementation of any funding advice

The ODX Breast Cancer Assay test is performed in a single laboratory in the United States by Genomic Health Inc. Therefore, the test would not be subject to approval or regulation by the Therapeutic Goods Administration (TGA). A November 2015 report by the US Food and Drug Administration (FDA) raised concerns about the current lack of regulation within the US for assays that are 'Laboratory Developed Tests' (LDTs), such as ODX.

MSAC previously raised concerns about the reliance on a single laboratory performing the test located in the US outside Australian standards maintained through the TGA or the National Association of Testing Authorities (NATA). MSAC also previously noted that a number of complex implementation issues would need to be considered by Government if this test was supported for listing in Australia.

6. Proposal for public funding

The proposal for public funding has changed since the previous resubmission (1342.4), and is presented in Table 2 (applicant highlighted changes with previous submission in red). The applicant has requested a fee of $5,085 per service, and the resubmission did not request any confidential pricing or fee arrangement.

Table 2 Proposal for public funding; changes from previous submission annotated (in red)

Gene expression profiling of tumour samples (surgical resection preferably or core biopsy) by reverse-transcriptase polymerase chain reaction (RT-PCR) technique for 21 genes in breast cancer tissue. *See Note for information on how results should be interpreted.* Previous submissions did not include a note on how results should be interpreted. May only be used to test samples from patients with all of the following characteristics as determined by the referring clinician: • early invasive breast cancer (stages I-II) No substantial change. • oestrogen receptor positive or progesterone receptor positive as determined by immunohistochemistry at an approved Australian pathology laboratory No substantial change. • HER2 negative as determined by immunohistochemistry and/or in situ hybridisation at an approved Australian pathology laboratory No substantial change. • node negative Previous submissions allowed for node positivity. Public funding no longer requested for node positive patients. • tumour size >= 10 mm and < 50 mm, or tumour size >= 5 mm and < 10 mm with unfavourable histological features (intermediate or poor nuclear and/or histologic grade, or lymphovascular invasion) The minimum tumour size of 2 mm has increased to 10 mm (or 5 mm with unfavourable histology). There was previously no maximum tumour size. Eligibility was also previously determined by the presence of 1 or 2 negative prognostic risk factors. • suitable for hormone therapy • suitable for adjuvant chemotherapy (ECOG performance status 0-2) • may only be used once per new primary breast cancer No substantial change. Fee: $5,085 Note: Chemotherapy decisions are guided by a patient's Recurrence Score (RS). Patients with RS<26 are recommended endocrine therapy and patients with RS≥26 are recommended adjuvant chemotherapy according to Oncotype DX. There is some evidence that there may be a chemotherapy benefit in patients aged ≤ 50 years, with RS 16-25. Previous submissions did not include a note on how results should be interpreted.

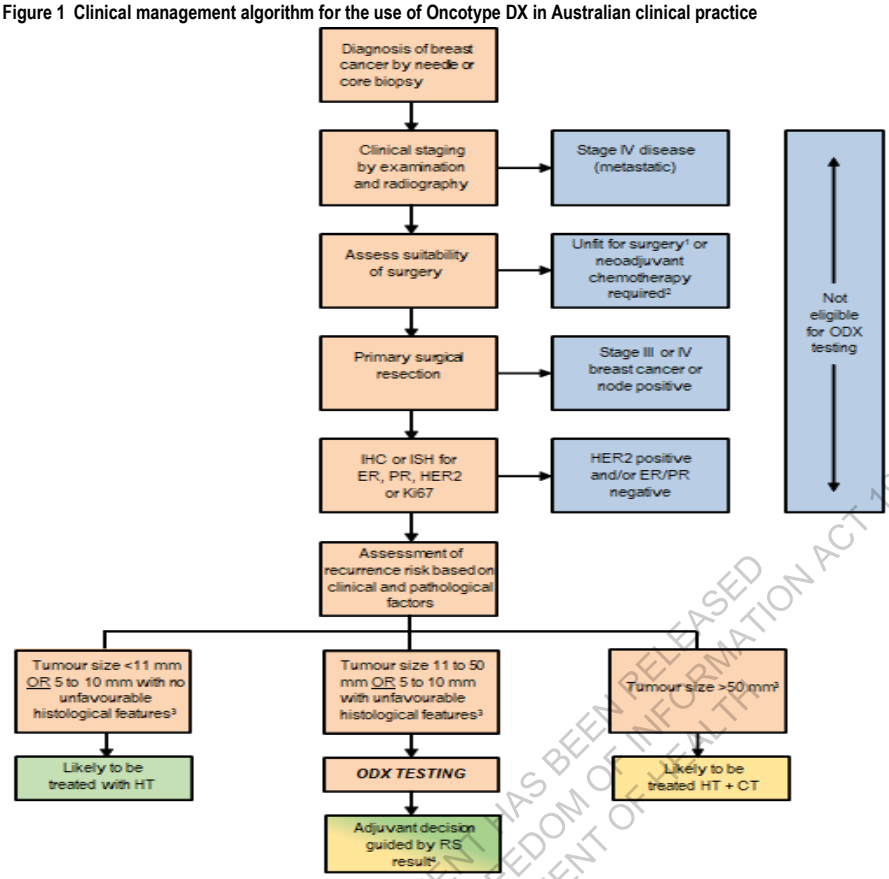
7. Summary of Public Consultation Feedback/Consumer Issues

See Application 1342.4 PSD on the MSAC website.

8. Proposed intervention's place in clinical management

The resubmission proposed clinical management algorithm (Figure 1) differs from that presented in earlier MSAC applications for Oncotype DX in that it excludes node positive patients, and the process used to exclude patients with very high or low clinical risk is based on the approach applied in TAILORx. In addition, the algorithm includes a footnote to clarify how recurrence score (RS) results should be interpreted and used to guide chemotherapy decisions.

Figure 1 Clinical management algorithm for the use of Oncotype DX in Australian clinical practice



[1] ODX is only appropriate for post-surgical patients.
[2] Patients who have received neo-adjuvant chemotherapy would continue with chemotherapy and Oncotype DX has not been validated for patients who have undergone neoadjuvant therapy
[3]Tumour size and grade parameters are based on eligibility for the TAILORx trial (Sparano, 2018)
[4]Chemotherapy decisions are guided by a patient's Recurrence Score (RS). Patients with RS<26 are recommended endocrine therapy and patients with RS≥26 are recommended adjuvant chemotherapy according to Oncotype DX; there is some evidence that there may be a chemotherapy benefit in patients aged ≤ 50 years, with RS 16-25.
Abbreviations: CT, chemotherapy; ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; HT, hormone therapy; IHC, immunohistochemistry; ISH, in situ hybridization; PR, progesterone receptor; ODX, Oncotype DX, RS, Recurrence Score

9. Comparator

The comparator for the current resubmission remains the same as that for the previous submissions - usual care. MSAC has previously accepted the comparator as usual care, defined as optimised subjective assessment of various clinical and pathological factors to estimate the risk of recurrence; which are likely combined using formal algorithms.

10. Comparative safety

The resubmission did not present a specific assessment of comparative safety. The Critique stated that the safety concerns remain as those outlined by MSAC previously and quoted in

the resubmission. "MSAC previously noted that although the test is procedurally safe because it relies on samples already taken for other purposes, there is a degree of risk in the misallocation of patients to risk categories, which would affect the outcomes of the therapy subsequently selected" (PSD for MSAC Application 1342, November 2013).

11. Comparative effectiveness

The resubmission is based on one prospective randomised trial and one re-analysis of a retrospective cohort study:

- The TAILORx trial was a prospective trial (N=10,273; registered population), that used a patient's recurrence score only to guide treatment. Women with intermediate RS (11-25) were randomised to endocrine therapy (ET) alone or ET+ chemotherapy (CT) (n=6,907; Arms B and C); and those with low (0-10; n=1,629; Arm A) or high (≥26; n=1,737; Arm D) RS were treated with ET alone or ET+CT, respectively (Sparano et al. NEJM, 2018). Results were provided for the 'main analysis set' or 'intention-to-treat (ITT) population' (n=9,719 across all four arms), and some results were also provided for the per protocol population ('as treated population'), which the Critique stated was an important comparison for demonstrating non-inferiority of ET alone vs. ET+CT. In addition, Sparano et al. stated comparisons of ITT population, stratified by randomisation, could still be biased because of differences in the group refusing chemotherapy (Arm C) and the group receiving chemotherapy (Arm B).
- Geyer et al. (2018) was a retrospective re-analysis of the NSABP B-20 trial (Fisher et al. 1997; Paik et al. 2006, previously considered by MSAC); a re-analysis of this study based on the recurrence scores used in the TAILORx trial and removing patients who were HER2-positive (Geyer et al. 2018).

TAILORx

The Critique presented forest plots for the primary outcome- invasive disease-free survival (iDFS) (Figure 2) and secondary outcome- freedom from recurrence at a distant site or distant recurrence-free interval (DRFI) (Figure 3).

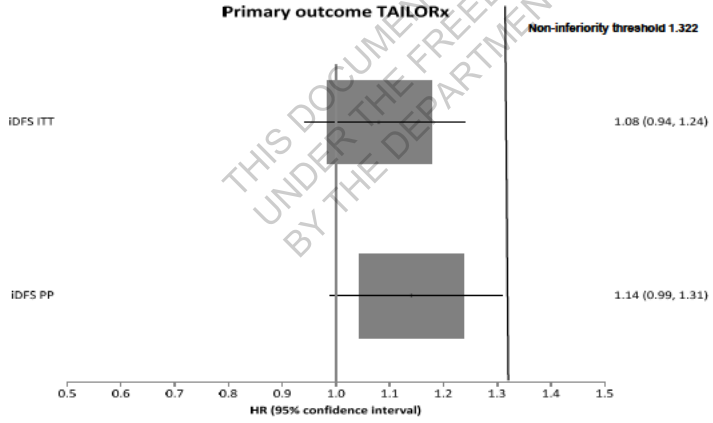


Figure 2 Forest plot of the hazard ratios (HR) of the intention-to-treat (ITT) and 'as-treated' (PP) populations, with the non-inferiority threshold for invasive disease-free survival (iDFS)

The primary analysis to support the claim of no difference between the treatment arms - endocrine therapy alone compared to endocrine therapy plus chemotherapy - met the prespecified non-inferiority threshold. However, the Critique outlined the following issues to consider:

- For the ITT population, the prespecified non-inferiority margin of 32.2% decrease in invasive disease-free survival for endocrine therapy alone compared to endocrine therapy plus chemotherapy appears to be quite large and not supported by the references cited in the trial report.
- Results for the 'as treated' population are close to rejecting the null hypothesis of no difference between the treatment arms.
- The 'as-treated' population baseline characteristics were statistically significantly different for important baseline prognostic variables such as age, menopausal status, tumour size and tumour grade (such that, on average, 'lower risk' women were randomised to ET alone and 'higher' risk women were randomised to ET+CT).
- The non-adherence to assigned therapy in the ET alone arm was 185/3458 = 5% but 608/3449 = 18% in the ET+CT arm, compared to only 89/1737 = 5% in the non-randomised high RS score chemotherapy arm.
- There was a high risk of bias in the trial design.
- There was significant loss to follow up which was deemed not important due to the lower than expected iDFS rate.
- There are four endocrine therapy regimens and nine chemotherapy regimens, which may introduce confounding to the extent that they are not equi-effective.

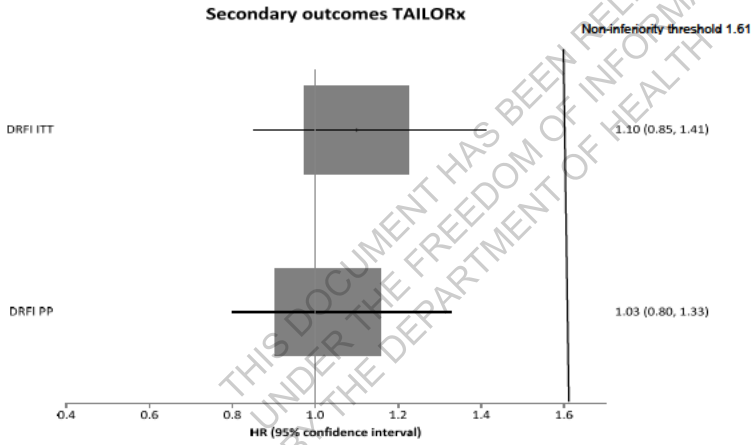


Figure 3 Forest plot of the hazard ratios (HR) of the intention-to-treat (ITT) and 'as treated' (PP) populations with non-inferiority threshold for distant recurrence-free interval (DRFI)

The secondary analysis to support the claim of no difference between the treatment arms - endocrine therapy alone compared to endocrine therapy plus chemotherapy - also met the prespecified non-inferiority threshold. However, the Critique outlined issues to consider:

- For the ITT population, the non-inferiority margin of a 61% decrease in freedom from recurrence at a distant site for endocrine therapy alone compared to endocrine therapy plus chemotherapy appears to be quite large and not supported by the references cited in the trial report.

- Full statistical power to do this comparison was not achieved: the prespecified number of events of 284 was not reached, but only 199 events were recorded.

Table 3 presents the estimated survival rates according to recurrence scores and assigned treatment in the ITT population. The Critique stated that similar issues as identified above for the primary and secondary analyses also occurred; the number of events required for full statistical power was not achieved and the evidence to support the assumptions for the prespecified non-inferiority threshold of 1.46 was not provided in the SBA or the trial report.

Table 3 Estimated survival rates according to RS and assigned treatment in the ITT population

End point and treatment group	Rate at 5 years (%)±SE	Rate at 9 years (%)±SE
Invasive disease-free survival		
Score of ≤10, endocrine therapy alone	94.0±0.6	84.0±1.3
Score of 11-25, endocrine therapy alone	92.8 ±0.5	83.3±0.9
Score of 11-25, chemotherapy + endocrine therapy	93.1±0.5	84.3±0.8
Score of ≥26, chemotherapy + endocrine therapy	87.6±1.0	75.7±2.2
Freedom from recurrence of breast cancer at a distant site		
Score of ≤10, endocrine therapy alone	99 3±0.2	96.8±0.7
Score of 11-25, endocrine therapy alone	98.0±0.3	94.5±0.5
Score of 11-25, chemotherapy + endocrine therapy	98 2±0.2	95.0±0.5
Score of ≥26, chemotherapy + endocrine therapy	93.0±0.8	86.8±1.7
Freedom from recurrence of breast cancer at a distant or local-regional site		
Score of ≤10, endocrine therapy alone	98.8±0.3	95.0±0.8
Score of 11-25, endocrine therapy alone	96 9±0.3	92.2±0.6
Score of 11-25, chemotherapy + endocrine therapy	97.0±0.3	92.9±0.6
Score of ≥26, chemotherapy + endocrine therapy	91.0±0.8	84.8±1.7
Overall survival		
Score of ≤10, endocrine therapy alone	98.0±0.4	93.7±0.8
Score of 11-25, endocrine therapy alone	98.0±0.2	93.9±0.5
Score of 11-25, chemotherapy + endocrine therapy	98.1±0.2	93.8±0.5
Score of ≥26, chemotherapy + endocrine therapy	95 9±0.6	89.3±1.4

Source: Table 7 of the Critique.

Geyer et al. (2018)

The re-analysis of the Paik et al. (2006) study by Geyer et al. (2018), considering only HER2-negative women and applying the 'old' and 'new' RS thresholds applicable for the definition of low, intermediate and high risk of recurrence is presented in Table 4. The Critique stated that the issues previously identified by MSAC about the 2006 Paik 2006 trial design remain.

Table 4 HR of adjuvant chemotherapy by RS subgroup, distant recurrence free survival (Geyer et al. 2018)

	N	Effect hazard ratio (95% CI)	P value
Overall (without HER2+ patients)	569	0 59 (0.31, 1.04)	Log rank P=0.06
Original RS subgroup n=569*	569		
Chemotherapy in RS <18	347	1.19 (0.40, 3.49)	
Chemotherapy in RS from 18-30	125	0.64 (0.23, 1.75)	
Chemotherapy in RS ≥31	97	0.18 (0.07, *0.46);*	
Likelihood ratio test on interaction			0.023
TAILORx RS groupings	569		
Chemotherapy in RS ≤10	176	1.19 (0.41, 3.51)	

	N	Effect hazard ratio (95% CI)	P value
Chemotherapy in RS 11-25	271	0.61 (0.26, 1 35)	
Chemotherapy in RS >25	122	0 27 (0.12, 0.62)	
Likelihood ratio test on interaction			0.014

Source: Tables 2 & 3 Geyer et al. 2018, Table 42 of the re-submission. Cox proportional Hazards Regression Model adjusted for patient age (>50 years vs ≤50 years), clinical tumour size (> 2.0 vs ≤2.0cm), ER by ligand blinding assay (≥100 vs <100 fmol/mg), PR by ligand blinding assay (≥100 vs <100 fmol/mg), and tumour grade (well differentiated, moderately differentiated and poorly differentiated.

Clinical claim

The Critique summarised the resubmission clinical claims:

- *A non-inferiority claim, for patients who the Oncotype DX test categorises into the intermediate recurrence group score, that endocrine therapy alone is no worse for the risk of distant recurrence free survival compared to endocrine therapy plus chemotherapy.*
- *A superiority claim, for patients who the Oncotype DX test categorises into the high recurrence group score, but usual care had determined treatment with endocrine therapy as sufficient, that the addition of chemotherapy would improve their disease free survival, risk of distant recurrence and overall survival.*

The non-inferiority claim is based on the results from TAILORx and the superiority claim is based on retrospective predictive data from the NSABP B-20 study (Paik et al. 2006; Geyer et al. 2018).

12. Economic evaluation

Table 5 summarises the economic evaluation.

Table 5 Summary of the economic evaluation

Perspective	Australian health care system
Comparator	Usual care, as defined by the M NDACT protocol used in TAILORx. Specifically, patients with low clinical risk do not receive adjuvant CT, patients with high clinical risk do receive adjuvant CT
Type of economic evaluation	Cost-utility analysis
Sources of evidence	TA LORx trial to determine allocation of CT in the usual care and Oncotype DX arms of the model NSABP B-20 Geyer et al. (2018) re-analysis to determine benefit of CT in patients who otherwise would not have received it
Time horizon	Lifetime
Outcomes	Life years gained, QALYs
Methods used to generate results	Markov cohort analysis
Health states	Free of disease recurrence • stratified by underlying Oncotype DX RS category and allocation to CT Disease recurrence Breast cancer death Other death
Cycle length	Annual
Discount rate	5% per annum
Software packages used	Microsoft Excel

The Critique stated that the model structure and modelling assumptions overwhelmingly favours Oncotype DX as all instances where Oncotype DX/RS score does not lead to optimal treatment were not considered, therefore the economic model presented is likely the most optimistic (and possibly implausible) scenario. The Critique presented the disaggregated incremental cost and effectiveness for "chemotherapy sparing" (Table 6) and "chemotherapy indicating" (Table 7) components of the model.

Table 6 Summary of disaggregated incremental cost and effectiveness in "chemotherapy sparing" only[a]

Parameter	Oncotype DX	Usual care	Incremental
Disaggregated costs			
Oncotype DX test costs	$5,085.00	$0.00	$5,085.00
Chemotherapy	$1,253.65	$3,116.03	-$1,862.38
Endocrine therapy	$3,160.85	$3,160.85	$0.00
Recurrent disease	$5,791.22	$5,791 22	$0.00
Total	**$15,290.72**	**$12,068.10**	**$3,222.62**
Disaggregated outcomes (discounted with half cycle correction)			
Life years	**13.6530**	**13.6530**	**0**
Disease-free	*13.4577*	*13.4577*	*0*
Post recurrence	*0.1953*	*0.1953*	*0*
QALY	**13.4621**	**13.4575**	**0.0045**
Disease-free	*13.3066*	*13.3021*	*0.0045*
Post recurrence	*0.1554*	*0.1554*	*0*
		$ per life year gained	**$NA**
		$ per QALY gained	**$711,529**

Text in italics indicate values calculated during the critique.
Source: 72 p155 of the SBA, ODX_EconModel xlsm.
[a] That is, moving any patients with RS ≤25 treated with ET+CT in the usual care arm to ET alone in the Oncotype DX arm.

Table 7 Summary of disaggregated incremental cost and effectiveness in "chemotherapy indicating" only[a]

Parameter	Oncotype DX	Usual care	Incremental
Disaggregated costs			
Oncotype DX test costs	$5,085.00	$0.00	$5,085.00
Chemotherapy	$3,672.22	$3,116.03	$556.19
Endocrine therapy	$3,175.34	$3,160.85	$14.50
Recurrent disease	$4,750.80	$5,791 22	-$1,040.43
Total	**$16,683.36**	**$12,068.10**	**$4,615.26**
Disaggregated outcomes (discounted with half cycle correction)			
Life years	**13.7665**	**13.6530**	**0.1135**
Disease-free	*13.6063*	*13.4577*	*0.1486*
Post recurrence	*0.1602*	*0.1953*	*-0.0351*
QALY	**13.5752**	**13.4575**	**0.1177**
Disease-free	*13.4466*	*13.3021*	*0.1445*
Post recurrence	*0.1275*	*0.1554*	*-0.0279*
		$ per life year gained	**$40,660**
		$ per QALY gained	**$39,217**

Text in italics indicate values calculated during the critique.
Source: 72 p155 of the SBA, ODX_EconModel xlsm.
[a] That is, moving any patients with RS ≥26 treated with ET alone in the usual care arm to ET+CT in the Oncotype DX arm.

The overall base case ICER is presented in Table 8 (combining the "chemotherapy sparing" and "chemotherapy indicating" components).

Table 8 Summary of disaggregated incremental cost and effectiveness from base case

Parameter	Oncotype DX	Usual care	Incremental
Disaggregated costs			
Oncotype DX test costs	$5,085.00	$0.00	$5,085.00
Chemotherapy	$1,809.84	$3,116.03	-$1,306.19
Endocrine therapy	$3,175.34	$3,160.85	$14.50
Recurrent disease	$4,750.80	$5,791 22	-$1,040.43
Total	**$14,820.98**	**$12,068.10**	**$2,752.88**
Disaggregated outcomes (discounted with half cycle correction)			
Life years	**13.7665**	**13.6530**	**0.1135**
Disease-free	*13.6063*	*13.4577*	*0.1486*
Post recurrence	*0.1602*	*0.1953*	*-0.0351*
QALY	**13.5798**	**13.4575**	**0.1222**
Disease-free	*13.4522*	*13.3021*	*0.1501*
Post recurrence	*0.1275*	*0.1554*	*-0.0279*
		$ per life year gained	**$24,253**
		$ per QALY gained	**$22,525**

Text in italics indicate values calculated during critique.
Source: Table 69, p153, Table 70 and 71 p154 of the SBA, ODX_EconModel xlsm.

The Critique highlighted that the base case ICER/QALY ($22,525) was driven by the "chemotherapy indicating" component (based on Geyer et al. 2018), contributing more benefit than the "chemotherapy sparing" component (incremental QALYs: 0.1177 *vs*. 0.0045, respectively); considered the "chemotherapy indicating" component was based on weaker evidence base, which MSAC had considered before when previously deciding not to support Oncotype DX.

The Critique's sensitivity analyses showed the modelled results were most sensitive to the effect of chemotherapy on absolute risk of recurrence in RS≥26 patients and the model duration.

13. Financial/budgetary impacts

An epidemiological approach has been used to estimate the financial implications of the introduction of the Oncotype DX test (Table 9).

Table 9 Net financial impact of Oncotype DX over five years by Commonwealth health budget and patient population

Summary	Year 1 (2020)	Year 2 (2021)	Year 3 (2022)	Year 4 (2023)	Year 5 (2024)
Patients diagnosed with breast cancer [A]	17,210	17,530	17,850	18,170	18,490
Number of patients eligible for Oncotype DX [B]	4,652	4,739	4,825	4,912	4,998
Number of patients using Oncotype DX testing [C]	1,396	1,896	2,171	2,456	2,749
Total expenditure on Oncotype DX [D]	$6,980,873	$9,480,899	$10,860,713	$12,283,795	$13,750,143
Critique values (removed $83.40 co-pay)	*$6,942,488*	*$9,428,768*	*$10,800,995*	*$12,216,251*	*$13,674,537*

Change in expenditure due to Oncotype DX [E]	-$1,795,774	-$2,438,885	-$2,793,832	-$3,159,908	-$3,537,114
Critique values (removed $83.40 co-pay)	*-$1,640,985*	*-$2,228,663*	*-$2,553,015*	*-$2,887,537*	*-$3,232,229*
Net impact of Oncotype DX on expenditure	**$5,185,099**	**$7,042,014**	**$8,066,882**	**$9,123,887**	**$10,213,029**
Critique values (removed $83.40 co-pay)	***$5,301,503***	***$7,200,104***	***$8,247,980***	***$9,328,715***	***$10,442,308***

[A] AIHW Cancer incidence projections; [B] 27% of [A]; [C] After applying expected uptake rates of 30 to 55%; [D] $5085 per test less patient contribution of $83.40 per test; [E] Savings of $1287 per patient tested due to reduction in chemotherapy.

The Critique stated that sensitivity analysis indicated that the estimates of net cost to the Commonwealth health budget is heavily reliant on the assumed uptake of the Oncotype DX test and also, but to a lesser extent, assumptions around cost offsets to the PBS.

14. Key issues from ESC for MSAC

ESC key issue	ESC advice to MSAC
Recurrence Score® (RS) thresholds for categorising low, intermediate and high risk of distant recurrence appear to be arbitrary and subject to change	The RS thresholds were modified in the context of the TAILORx trial. It is not unreasonable to adjust parameters based on additional data, and the new threshold level of 26 appears safe based on the TA LORx and other supporting studies.
Population (as per the eligibility criteria into the TA LORx trial)	The eligible population should be specified as patients with newly diagnosed breast carcinomas; who are *ER*-positive, *HER2*-negative, lymph node-negative and post-surgical; and who have not received neoadjuvant therapy.
Proposed note defining eligibility for funding should be modified, as it suggests that patients with an RS ≥26 should receive chemotherapy only	TA LORx trial protocol specified that women with an RS score of ≥26 were assigned to receive chemotherapy plus endocrine therapy. Therefore, this should be reflected in the note.
Clinical need	There is a view among clinicians that knowledge of the genomic features of breast cancers is required to provide a higher level of evidence on which to base systemic treatment decisions. Multigene assays are being employed routinely by clinicians in the US.
Context	Oncotype DX represents one of the more rigorously developed gene assays with good quality control; NCCN preferred and 'strong' recommendation by ASCO.
Uncertain chemotherapy benefit – 26% or 15% or 20 5%?	20 5% may be an acceptable estimate.
Costs of adding chemotherapy may be underestimated	The cost of chemotherapy needs to be revisited – if it is higher, cost offsets would be higher.
Test is not registered for use in Australia and a single laboratory in the US performs the test and may not be eligible for listing on the MBS. Who will pay for this? What about out-of-pocket costs?	Since testing is done outside Australia, is it possible for MBS to pay the small pathology fee for collecting and preparing the sample to be sent, and then adopt a separate arrangement to reimburse the patient for the rest?
Different results from economic model depending on accepting different sources of clinical evidence	Given MSAC's published views on the strength of the evidence available previously, it may be useful for MSAC to consider the disaggregated analyses of the non-inferiority (based on TA LORx) and effectiveness (based on re-analysing the previous retrospective predictive evidence) components of the model.

ESC discussion

Application 1342.5 is a resubmission seeking public funding for a gene expression profiling test, Oncotype DX®, for patients with breast cancer. The test generates a Recurrence Score®

(RS) that is used to predict the likelihood of breast cancer recurrence and the potential benefit of also receiving adjuvant chemotherapy for surgically treated patients with early-stage invasive breast cancer receiving adjuvant endocrine therapy.

ESC noted the resubmission includes two therapeutic claims:

1. Oncotype DX will identify patients who would not benefit from also receiving adjuvant chemotherapy, thus sparing them the adverse effects and other risks associated with chemotherapy (referred to as "chemotherapy sparing"; RS <26)

2. Oncotype DX will identify patients likely to benefit from also receiving adjuvant chemotherapy who would not have been identified through standard clinical practice; appropriate use of chemotherapy will result in improved disease-free survival (referred to as "chemotherapy indicating"; RS ≥26).

ESC noted MSAC's previous concerns about reliance on a single United States (US) laboratory performing the test. However, ESC considered that centralisation of testing could be seen as a significant strength of Oncotype DX in terms of reproducibility. It does not suffer from the same problems as other assays based on technologies that are difficult to standardise across different laboratories. Hence, there is no laboratory-based need for an Australian laboratory to implement new testing strategies.

ESC noted that the US Food and Drug Administration is currently obtaining guidance and feedback on its proposed oversight of laboratory-developed tests such as Oncotype DX, but new guidelines are not yet in place. The laboratory is accredited by the College of American Pathologists under the US Clinical Laboratory Improvement Amendment (CLIA) of 1988, which has parallels with accreditation by the National Association of Testing Authorities (NATA) in Australia.

ESC noted that the resubmission used the structure of an MBS item with descriptor, fee and note to frame its request for public funding. The note is intended to help interpret RS scores for making chemotherapy decisions. It states that patients with RS<26 are recommended endocrine therapy and patients with RS≥26 are recommended adjuvant chemotherapy. However, ESC noted that the TAILORx trial protocol specified that women with a score of ≥26 were assigned to receive adjuvant chemotherapy plus endocrine therapy. This should be reflected in the note.

ESC noted that the proposed fee of $5,085 per test service is higher than the confidential fee in previous submissions ($3,375). The applicant has proposed that $85 of the fee is for the Australian pathology laboratory retrieving and preparing the tissue.

ESC noted that some of the PICO criteria have changed since the previous MSAC considerations of this application, to align with the TAILORx trial:

- population – narrowed to include node negative-women with larger tumour size (the initial submission and first resubmission allowed for node positivity, while the second and third resubmissions excluded lymph node positivity but allowed smaller tumour sizes)
- intervention – RS threshold for decision-making with respect to recommending adjuvant chemotherapy as well as receiving adjuvant endocrine therapy is now 26 instead of 31
- comparator – usual care is now more clearly defined, and aligned with the MINDACT protocol used in TAILORx.

ESC considered that the eligible population should be specified as patients with newly diagnosed breast carcinomas *ER+*, *HER2–*, lymph node-negative who are post-surgical and who have not received neoadjuvant therapy. Restrictions might also include requesting by a specialist medical or surgical oncologist.

Although changing the RS threshold will change the consequences for the eligible population, ESC noted that the TAILORx trial was specifically designed to establish whether treating women with a mid-range RS of 11–25 with adjuvant endocrine therapy alone results in significantly worse breast cancer outcomes compared treating these women with both adjuvant chemotherapy and adjuvant endocrine therapy. This is the patient group for whom the decision around the use of adjuvant chemotherapy is not clear based on clinical–pathological factors such as tumour size and grade.

From the consumer point of view, ESC noted that genomics is becoming a part of better patient-centred care. There is considerable positive benefit for patients of better diagnoses leading to better treatment decisions, including patients being able to avoid chemotherapy if it is not required. ESC noted that equity of access issues arise from this test not being rendered in Australia.

ESC noted that Oncotype DX is a rigorously developed gene assay with good quality control. It is given a 'strong' recommendation in the American Society of Clinical Oncology (ASCO) guidelines, and the National Comprehensive Cancer Network (NCCN) has designated it as the preferred multigene panel assay.

ESC noted that other countries fund Oncotype DX. The National Institute for Health and Care Excellence (NICE) recommended it in 2013 for coverage under the England's National Health Service (NHS), for use in early-stage *ER+*, *HER2–*, node-negative invasive breast cancer patients with 'intermediate risk'. Coverage was renewed in 2018 and expanded to include patients with micrometastases. Node-positive disease is not yet covered by the NHS, but some patients are covered by private insurance.

Oncotype DX is publicly funded for almost all eligible patients in England, with no patient co-payment. Genomic Health Inc. estimates that 95% of the trusts serving breast cancer patients in the UK use the test, and over 22,000 women in the UK had undergone the test as of late 2018.

In Canada, all 10 provinces provide Oncotype DX under their public healthcare systems. Seven of the 10 provinces provide the test for node-negative and micrometastases patients; three provinces also provide, and one is considering providing, the test for node-positive patients.

In the USA, Oncotype DX is covered by Medicare (which covers people over 65 years of age) in all states except two, and by Medicaid (which covers people on low incomes) in all 50 states. The test is also covered by all major private insurers. Medicare and other public systems cover node-negative and node-positive patients; about half the private insurers cover node-positive patients.

ESC noted that there is an increasing view that clinicians should be using a higher level of evidence based on genomic subtyping of individual cancers (in addition to traditional histological features and immunohistochemical markers) to provide more specific and tailored treatments for breast cancer patients. Oncotype DX and other similar multigene assays are being increasingly used worldwide, and there is an increasing clinician-led demand

for access to these types of assays. Assays like Oncotype DX are intended for use as an additional tool to guide decision-making, not to dictate treatment. ESC noted that clinicians and researchers are also currently using whole exome sequencing (WES) and whole genome sequencing (WGS) to investigate the genomic profile of breast cancers.

ESC considered that most clinicians would order the Oncotype DX assay selectively, particularly in instances when decision-making is complex. However, ESC considered that there is some risk of leakage. ESC noted that NICE guidance for Oncotype DX has recently been updated, which may inform concerns regarding leakage.

ESC noted the limitations of the current online prediction tools used to estimate the risk of recurrence and to make treatment decisions (Wazir et al. 2017):

- Adjuvant! Online tends to overestimate the number of patients at high risk; overestimate the survival rates of younger women with ER-positive breast cancer; overestimate the added value of chemotherapy for older patients; and *HER2* assessment is not included

- NHS Predict does not provide any estimate of local relapse; and does not consider mortality due to causes other than breast cancer. Some patients, particularly those with small, biologically aggressive cancers, may therefore not receive chemotherapy that would be of benefit.

ESC noted that the previously provided retrospective predictive data from the randomised NSABP B-20 study (Paik et al. 2006) is again relied on to support the clinical claim that Oncotype DX will identify patients likely to benefit from also receiving adjuvant chemotherapy who would not have been identified through standard clinical practice. The re-analysis of these data by Geyer et al. 2018 is relied on to demonstrate that also receiving adjuvant chemotherapy is superior to endocrine therapy alone in patients with RS ≥26.

ESC noted that the TAILORx trial provides NHMRC Level II evidence that adjuvant chemotherapy can be withheld in patients with an RS ≤26 without affecting the patient's risk of disease recurrence (Sparano et al. 2018). ESC also noted that exploratory analyses indicated that also receiving adjuvant chemotherapy was associated with some benefit for women aged ≤50 years with an RS of 16–25.

ESC noted that two Australian Decision Impact Studies (ADIS) previously presented to MSAC are used in the resubmission to characterise current patterns of care. These data are used to investigate the applicability of usual care in TAILORx to Australian practice. One of these studies (de Boer et al. 2013) found that the Oncotype DX RS changed the treatment recommendation in 24% of patients with node-negative tumours. In the other study (Chin-Lenn et al. 2018), the Oncotype DX RS changed treatment recommendations in 38% of patients, noting that the change in treatment recommendation could be in either direction: to include chemotherapy when it would have otherwise been excluded, or to exclude chemotherapy when it would otherwise have been included. However, ESC considered that the lack of proven clinical utility in the Australian context to be an ongoing issue. There is still no good description of current Australian practice as the ADIS studies are now several years old. It is likely to be different to practice in the US and UK, and it cannot be assumed that incremental clinical utility will be the same in Australia as in other countries.

ESC noted that the cost of adjuvant chemotherapy used in the model revised since the previous submission was recalculated by the applicant using the Critique's assumption of four cycles rather than six. However, ESC noted the applicant's comment in response that the

revised cost is likely to be an underestimate of the true burden of this chemotherapy to the health care system. ESC commented that most adjuvant chemotherapy treatments go beyond four cycles so the cost might be underestimated, and noted that if this cost is higher, cost offsets would be higher.

ESC noted that the period of adjuvant chemotherapy treatment was based on six cycles; the applicant agreed to base this cost on four cycles but did not change the disutility duration to reflect four cycles. ESC queried whether using four cycles would reduce the estimate of quality-adjusted life years gained from avoiding the toxicity of adjuvant chemotherapy.

ESC noted translation issues arising from uncertainty regarding the appropriate extent of benefit (i.e. reduction in absolute risk of disease recurrence) of receiving adjuvant chemotherapy as well as adjuvant endocrine therapy in patients with an RS ≥26. The applicant originally used a value of 26% (based on Geyer et al.), but the Critique suggested 15% would be more appropriate in the Australian context. Instead, the applicant reduced the incremental benefit of chemotherapy from 26% in the base case to a mid-point of 20.5%. ESC advised that 20.5% may be acceptable.

ESC noted that the revised model uses revised utility values, which are now more in line with TAILORx.

ESC noted that the base case ICER/QALY from the revised combined model is sensitive to several assumptions, which vary this estimate within the range of $22,000–$50,000 (using a chemotherapy benefit of 20.5%). However, ESC noted that the ICER/QALY calculated using a chemotherapy benefit of 15% was more than $67,500.

ESC noted that although the economic evaluation model is correct, it is basic. It includes only univariate sensitivity analyses, but no probability sensitivity analysis or cost-effectiveness acceptability curve. The model includes direct costs only; it does not include out-of-pocket costs. ESC queried whether the PBS cost of new chemotherapy drugs used in the TAILORx trial had been included in the cost offsets.

ESC noted that the analysis also gave two results based on the source of clinical utility evidence: evidence for the non-inferiority claim is from the TAILORx randomised trial, but the economic analysis is driven by superiority claim from the retrospective predictive re-analysis from Paik/Geyer. ESC noted that it may be useful for MSAC to consider the disaggregated analyses of the non-inferiority and superiority components of the model (as well as the combined analysis).

ESC noted that the applicant's revised financial analyses resulted in a modest increase in the net budgetary impact to $44.7 million over the first 5 years. The applicant also provided a revised estimate incorporating updated (2017) breast cancer incidence data from the Australian Institute of Health Welfare of $50.3 million over the first 5 years. ESC considered these two estimates to be more realistic than the estimate of $51.6 million over 5 years using UK uptake data. However, ESC considered that the financial estimates remained subject to significant uncertainty due to low uptake rate assumptions and the fact that the TAILORx trial did not report important patient baseline characteristics, such as the percentage expression of ER or PR.

15. Other significant factors

Nil.

16. Applicant's comments on MSAC's Public Summary Document

The MSAC Executive 3 February 2012 teleconference agreed for MSAC applicants to be given the opportunity to have a comment inserted in the final outcomes document – to be limited to one paragraph and/or a link to reference material

17. Further information on MSAC

MSAC Terms of Reference and other information are available on the MSAC Website: visit the MSAC website

s 47C, s 47E

s 47C, s 47E

s 47C, s 47E

The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a retrospective analysis

Joseph C Del Paggio, Ian F Tannock

Summary

Background The fragility index of trial results—ie, the minimum number of changes from non-events to events resulting in loss of statistical significance—can provide a measure of confidence that a positive effect reported in a randomised controlled trial is real. We aimed to calculate the fragility index of randomised controlled trials supporting US Food and Drug Administration (FDA)-approved anticancer drugs.

Methods This is a retrospective analysis of phase 3, randomised, controlled trials supporting anticancer drugs that were approved by the FDA between Jan 1, 2014, and Dec 31, 2018. Two-arm studies with 1:1 randomisation and significant positive results for a time-to-event outcome were eligible for the fragility index calculation, which involves the iterative addition of an event to the experimental group (defined as the group with the smaller number of events in positive trials) and concomitant subtraction of a non-event from that group, until positive significance (defined as p<0·05 by Fisher's exact test) is lost.

Findings We identified 36 phase 3 randomised controlled trials, of which 17 (47%) were included in the fragility index analysis. The median fragility index was 2 (IQR 0–27). The fragility index was 2 or less in nine (53%) of 17 trials; for these trials, the fragility index was 1% or less of the total sample size. In five (29%) of 17 trials, the number lost to follow-up was more than the fragility index.

Interpretation Many phase 3 randomised controlled trials supporting FDA-approved anticancer drugs have a low fragility index, challenging confidence for concluding their superiority over control treatments. Although not a measure of effect, the fragility index might provide an additional means of assessing the robustness of clinical trial data.

Funding None.

Lancet Oncol 2019

Published Online
July 8, 2019
http://dx.doi.org/10.1016/S1470-2045(19)30338-9

Department of Medical Oncology, Thunder Bay Regional Health Sciences Centre and Northern Ontario School of Medicine, Thunder Bay, ON, Canada (J C Del Paggio MD); and Division of Medical Oncology, Princess Margaret Cancer Centre and University of Toronto, Toronto, ON, Canada (Prof I F Tannock MD)

Correspondence to:
Prof Ian F Tannock, Division of Medical Oncology, Princess Margaret Cancer Centre, Toronto, ON M5G 2X9, Canada
ian.tannock@uhn.ca

Introduction

Reliance on p values for establishing significance of results in clinical trials is contentious.[1] Outcomes meeting the arbitrary threshold of a p value less than 0·05 might not be clinically relevant, particularly when the difference in outcome does not provide substantial clinical benefit according to European Society for Medical Oncology or American Society of Clinical Oncology value scales.[2,3]

The statistical fragility of the results of randomised controlled trials can be represented by the ease with which its threshold p value shifts from significant (p<0·05) to non-significant (p≥0·05) when experimental outcomes change from non-events to events. Walsh and colleagues[4] defined the fragility index as the minimum number of such changes, and the fragility index provides a measure of confidence that a positive effect reported in a randomised controlled trial comparing an experimental to a control treatment is real. The purpose of this study is to assess the fragility of phase 3 trials supporting recent US Food and Drug Administration (FDA)-approved anticancer drugs.

Methods

Study design

We reviewed FDA approvals for anticancer medicines between Jan 1, 2014 and Dec 31, 2018, publicly available at the FDA website. Using Google Scholar, we identified the phase 3 trials supporting each drug for the indication for which it was approved. For the fragility index analysis, we included only two-arm studies with 1:1 randomisation that reported significant positive primary outcome results for a time-to-event outcome[4] for the intention-to-treat population; secondary endpoints were assessed in cases in which the primary endpoint was not significant (ie, p≥0·05 or the upper limit of the CI crossed 1). We abstracted information on trial design, observed numbers of events for the control and experimental groups for primary or secondary time-to-event outcomes, and the number of patients lost to follow-up. Data not available in the primary publication or its appendix were augmented by data in ClinicalTrials.gov or in Statistical Review and Evaluation documents on the FDA website.

The fragility index was calculated from a two by two contingency table by the iterative addition of an event to the experimental group (defined as the group with the smaller number of events in positive trials) and concomitant subtraction of a non-event from that same group, thereby maintaining a constant total number of events plus non-events, until positive significance (defined as p<0·05) was lost. p values were calculated

For the **FDA website**, see https://www.fda.gov

For **ClinicalTrials.gov**, see https://clinicaltrials.gov/

Research in context

Evidence before this study

We searched Google Scholar with the search terms "fragility index" and "oncology" or "FDA" for estimates of trial fragility in anticancer trials. Studies in areas of medicine other than oncology have shown that a change in a small number of events from negative to positive in the experimental group of a randomised controlled trial can lead to loss of significance; the number of events to provide loss of significance has been termed the fragility index. We found no previous estimates of trial fragility for randomised trials assessing anticancer drugs.

Added value of this study

We did a review of the FDA website to identify anticancer drugs approved in the 5-year interval of 2014–18. We identified two-arm phase 3 trials reporting significant positive results that supported FDA approval of an anticancer drug. We calculated the fragility index for each of these trials. To our knowledge, this is the first assessment of fragility index for randomised controlled trials in oncology.

Implications of all the available evidence

We found that many phase 3, randomised, controlled trials supporting FDA-approved anticancer drugs have a low fragility index, often less than 1% of the sample size or less than the number of patients lost to follow-up. Our results show that approval of many anticancer drugs is based on fragile evidence. The fragility index might provide an additional means of assessing the robustness of clinical trial data.

For the **online fragility index calculator** see https://clincalc.com/Stats/FragilityIndex.aspx

with Fisher's Exact Test.[4] An online fragility index calculator is available.

Role of the funding source

There was no funding source for this study. Both authors had full access to all data used in the study. The corresponding author had final responsibility for the decision to submit for publication.

Results

Between Jan 1, 2014, and Dec 31, 2018, the FDA approved 55 anticancer drugs for novel indications, of which 32 (58%) were for solid tumours. 29 (53%) of the drugs were approved on the basis of phase 1 or phase 2 trials. Five (9%) of the 55 approvals were based on the endpoint of overall survival, 19 (35%) were based on an alternative time-to-event endpoint, most often progression-free survival, 30 (55%) were based on measures of tumour response, and one (2%) was based on pharmacokinetic data. Phase 3 trials for the approved settings or indications were identified for 36 (65%) of 55 drugs.

17 (47%) of 36 phase 3 randomised controlled trials met the inclusion criteria for fragility index analysis (table); we could not calculate the fragility index for the remaining 19 phase 3 trials because of unequal allocation between groups (16 trials) or a statistically negative time-to-event endpoint (three trials). The median sample size for the 17 eligible randomised controlled trials was 452 (range 220–2840). The primary endpoint was used for fragility index analysis in 16 of the 17 eligible trials;[5–16,18–21] for the remaining trial,[17] the fragility index was calculated using the secondary progression-free survival endpoint, because the primary overall survival endpoint was not significant (table).

The median fragility index for the 17 studies was 2 (IQR 0–27)—ie, a median of two events was required to change the results of the endpoint analysis from significant to non-significant (figure). Fragility index was 2 or less in nine (53%) of 17 trials;[5,8–11,15,17,19,20] for these trials, the fragility index was 1% or less of the total sample size. For the six trials with a fragility index of 0 (ie, Fisher's exact test p>0·05), the χ^2 test (one trial[9]) and stratified log-rank test (five trials[5,10,15,17,19]) had been used to calculate the reported significant p value. In six (35%) of 17 trials,[9,10,12,16,17,21] the number of patients lost to follow-up was two or more (median 1 [IQR 0–2; range 0–68]); the number lost to follow-up was more than the respective fragility index in five (29%) of the 17 trials.[9,10,16,17,19] Of the 17 drugs tested in the eligible trials included in the fragility index analysis, only one drug (daratumumab) was supported by more than one positive phase 3 trial in that setting and indication (table).[13,22]

Discussion

In our retrospective analysis, we show that about half of the phase 3 trials supporting FDA-approved anticancer drugs have a low fragility index and are vulnerable to losing significance with a change in designation of very few events, often a change in event number less than 1% of the respective trial sample size. The change in number of events required for fragility is also often smaller than the number of patients lost to follow-up, raising concerns about a statistical change in the results had these patients been assessed to their endpoints.

To our knowledge, no previous studies have estimated the fragility index for oncology trials or for trials strictly supporting FDA-approved medications. The fragility index has been applied to other randomised controlled trials, including those assessing spinal surgery,[23] critical care,[24] and heart failure,[25] and to trials supporting clinical practice guidelines.[26,27] These studies are consistent in showing that many randomised controlled trials are fragile, and several investigators have recommended adoption of the fragility index in reporting clinical trial outcomes.[24,27,28] Trials with large fragility indexes are present in our cohort; however, most trials were powered to detect differences in progression-free survival, which is subject to biases of clinical and radiological assessment, as well as informative

	Approval date	Disease site or indication	Trial	Endpoint	Experimental sample size	Experimental event number*	Control sample size	Control event number*	p value†	Fragility index
Ceritinib	April 2014	Non-small-cell lung cancer	ASCEND-5[5]	Progression-free survival	115	83	116	89	<0·0001	0
Idelalisib	July 2014	Chronic lymphocytic leukaemia	GS-US-312-0116[6]	Progression-free survival	110	12	110	53	<0·001	26
Panobinostat	February 2015	Multiple myeloma	PANORAMA1[7]	Progression-free survival	387	207	381	260	<0·0001	31
Dinutuximab	March 2015	Neuroblastoma	ANBL0032[8]	Event-free survival	113	33	113	50	0·0115	2
Elotuzumab	November 2015	Multiple myeloma	ELOQUENT-2[9]	Progression-free survival	321	179	325	205	<0·001	0
Necitumumab	November 2015	Squamous non-small-cell lung cancer	SQUIRE[10]	Overall survival	545	418	548	442	0·012	0
Ixazomib	November 2015	Multiple myeloma	TOURMALINE-MM1[11]	Progression-free survival	360	129	362	157	0·012	2
Cobimetinib	November 2015	Melanoma	coBRIM[12]	Progression-free survival	247	79	248	128	<0·001	27
Daratumumab	November 2015	Multiple myeloma	CASTOR‡[13]	Progression-free survival	251	67	247	122	<0·0001	35
Venetoclax	April 2016	Chronic lymphocytic leukaemia	MURANO[14]	Progression-free survival	194	32	195	114	<0·001	62
Midostaurin	April 2017	Acute myeloid leukaemia	RATIFY[15]	Overall survival	360	171	357	186	0·002	0
Neratinib	July 2017	Breast cancer	ExteNET[16]	Invasive disease-free survival	1420	70	1420	109	0·0091	12
Inotuzumab ozogamicin	August 2017	Acute lymphocytic leukaemia	INO-VATE ALL[17]	Progression-free survival§	164	129	162	128	<0·001	0
Lutetium (^{177}Lu) oxodotreotide	January 2018	Gastroenteropancreatic neuroendocrine tumours	NETTER-1[18]	Progression-free survival	116	23	113	68	0·004	32
Mogamulizumab	August 2018	Mycosis fungoides or Sézary syndrome	MAVORIC[19]	Progression-free survival	186	110	186	122	<0·0007	0
Duvelisib	September 2018	Chronic lymphocytic leukaemia	DUO[20]	Progression-free survival	160	93	159	110	<0·0001	1
Dacomitinib	September 2018	Non-small-cell lung cancer	ARCHER 1050[21]	Progression-free survival	227	136	225	179	<0·0001	27

FDA=US Food and Drug Administration. *Event numbers established by independent review committees were abstracted preferentially when available. †Calculated by the statistical methods in each trial. ‡A second phase 3 trial, POLLUX,[22] was published in the same treatment setting; it has a progression-free survival fragility index of 42. §Significant secondary endpoint.

Table: **Fragility index calculated for 17 phase 3 trials with 1:1 randomisation supporting drugs approved by the FDA between 2014 and 2018**

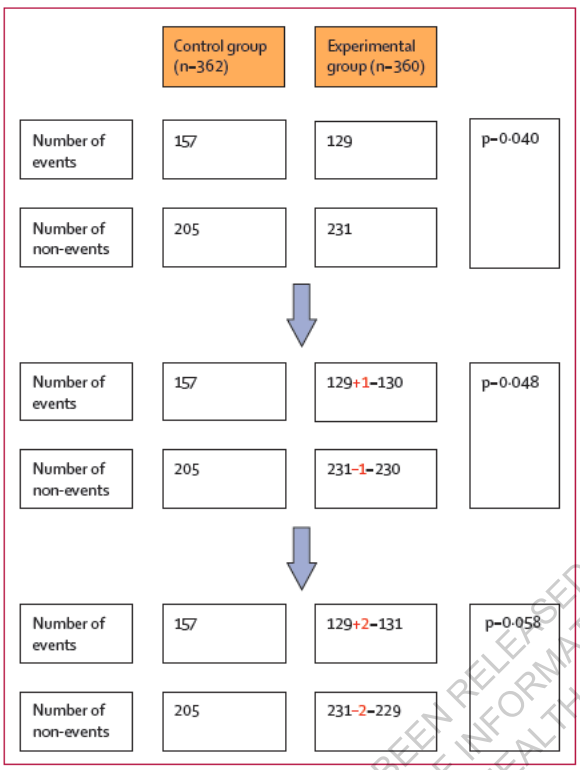


Figure: **Example of fragility index calculation for the phase 3 trial TOURMALINE-MM1[11]**
p values are calculated by Fisher's exact test, whereas the p value in the original study was calculated as 0·012 using the stratified log-rank test. The fragility index in this example is 2, which is the number of non-events required to convert to events so that the difference between the control and experimental groups no longer meets significance at the α=0·05 level using Fisher's exact test.

censoring, where there is loss of patients to follow-up before meeting criteria of progression.[29] A low fragility index in trials using progression-free survival or similar endpoints might be one of several factors leading to poor correlation with overall survival.

In principle, the p value is an indication of the compatibility between data from a trial and the prespecified statistical model: smaller p values imply greater statistical incompatibility of the data with the null hypothesis—a postulate of no difference between outcomes of the experimental and control group.[30] The p value depends on assumptions. The log-rank test used in survival analysis has the advantage that it accounts for events over time, but it relies on the assumption that the hazard ratio of two treatments is constant over time (ie, proportional hazards). Fisher's exact test (used to calculate the fragility index) has the disadvantage that it does not account for the time at which events occurred,[31] but it does not require proportional hazards, a condition that is not satisfied, for example, when survival curves cross. Fragility index calculations of zero (indicting a p≥0·05 by Fisher's exact test) were possible for trials reporting significance based on log-rank and other tests: these indicate extreme fragility.

The approval by the FDA of anticancer drugs considers the totality of evidence relating to their effectiveness in the context of the illness for which the drug is intended, the risks of the drug and management thereof, the uncertainties in extrapolating clinical data to the real world, and the applicable laws and regulations. Clinical data from phase 2 trials as well as phase 3, randomised, controlled trials might be analysed in the approval process; however, results of phase 2 trials can be misleading, and phase 3, randomised, controlled trials are regarded as providing the highest level of evidence relating to clinical benefit. For the drugs approved in the 5-year period under analysis in this study, it was rare that approval was supported by more than one phase 3, randomised, controlled trial.

This study is limited by its small sample size, necessitated by the number of oncological drugs approved by the FDA within the study period, as well as the 1:1 randomisation required for the fragility index calculation.[4] The operating characteristics of the fragility index also limit its use in time-to-event data: in situations where the number of events is similar between two groups, but a difference in timing exists, the fragility index might be overly sensitive in concluding fragility.[4] Finally, since a strong relationship exists between the p value and the fragility index,[32] caution must be taken in concluding the robustness of a clinical trial on fragility index alone without a broader context (eg, statistical design, effect size, CIs, and minimal important differences). As exemplified by the outcome data we presented, extreme fragility can be noted in situations in which the absolute difference in numbers of events between the experimental and control groups is quite large—a difference that might be considered clinically meaningful. In general, however, larger overall sample sizes increase the fragility index,[28] which speaks to the aforementioned correlation between the fragility index and p value.

The finding that many phase 3, randomised, controlled trials supporting FDA-approved anticancer drugs have a low fragility index challenges the confidence in concluding superiority for these drugs over control treatments. Many FDA-approved drugs have been shown to be of low clinical value,[33,34] and measuring the robustness of clinical trial data to support their high cost is paramount. The fragility index, like the p value, should not be interpreted as a measure of effect, but it can shed some light on the strength of statistical conclusions.

Contributors
All authors contributed to the design of the study. JCD acquired the data for the study, performed the analyses, and drafted the report. All authors interpreted the results and contributed to the writing of the final report.

Declaration of interests
We declare no competing interests.

References
1 Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019; 567: 305–07.
2 Cherny NI, Dafni U, Bogaerts J, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Ann Oncol* 2017; 28: 2340–66.

3 Schnipper LE, Davidson NE, Wollins DS, et al. Updating the American Society of Clinical Oncology value framework: revisions and reflections in response to comments received. *J Clin Oncol* 2016; **34:** 2925–34.

4 Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. *J Clin Epidemiol* 2014; **67:** 622–28.

5 Shaw AT, Kim TM, Crinò L, et al. Ceritinib versus chemotherapy in patients with ALK-rearranged non-small-cell lung cancer previously given chemotherapy and crizotinib (ASCEND-5): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* 2017; **18:** 874–86.

6 Furman RR, Sharman JP, Coutre SE, et al. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N Engl J Med* 2014; **370:** 997–1007.

7 San-Miguel JF, Hungria VT, Yoon SS, et al. Panobinostat plus bortezomib and dexamethasone versus placebo plus bortezomib and dexamethasone in patients with relapsed or relapsed and refractory multiple myeloma: a multicentre, randomised, double-blind phase 3 trial. *Lancet Oncol* 2014; **15:** 1195–206.

8 Yu AL, Gilman AL, Ozkaynak MF, et al. Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. *N Engl J Med* 2010; **363:** 1324–34.

9 Lonial S, Dimopoulos M, Palumbo A, et al. Elotuzumab therapy for relapsed or refractory multiple myeloma. *N Engl J Med* 2015; **373:** 621–31.

10 Thatcher N, Hirsch FR, Luft AV, et al. Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controlled phase 3 trial. *Lancet Oncol* 2015; **16:** 763–74.

11 Moreau P, Masszi T, Grzasko N, et al. Oral ixazomib, lenalidomide, and dexamethasone for multiple myeloma. *N Engl J Med* 2016; **374:** 1621–34.

12 Ascierto PA, McArthur GA, Dréno B, et al. Cobimetinib combined with vemurafenib in advanced BRAF(V600)-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. *Lancet Oncol* 2016; **17:** 1248–60.

13 Palumbo A, Chanan-Khan A, Weisel K, et al. Daratumumab, bortezomib, and dexamethasone for multiple myeloma. *N Engl J Med* 2016; **375:** 754–66.

14 Seymour JF, Kipps TJ, Eichhorst B, et al. Venetoclax-rituximab in relapsed or refractory chronic lymphocytic leukemia. *N Engl J Med* 2018; **378:** 1107–20.

15 Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N Engl J Med* 2017; **377:** 454–64.

16 Chan A, Delaloge S, Holmes FA, et al. Neratinib after trastuzumab-based adjuvant therapy in patients with HER2-positive breast cancer (ExteNET): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* 2016; **17:** 367–77.

17 Kantarjian HM, DeAngelo DJ, Stelljes M, et al. Inotuzumab ozogamicin versus standard therapy for acute lymphoblastic leukemia. *N Engl J Med* 2016; **375:** 740–53.

18 Strosberg J, El-Haddad G, Wolin E, et al. Phase 3 trial of ^{177}Lu-dotatate for midgut neuroendocrine tumors. *N Engl J Med* 2017; **376:** 125–35.

19 Kim YH, Bagot M, Pinter-Brown L, et al. Mogamulizumab versus vorinostat in previously treated cutaneous T-cell lymphoma (MAVORIC): an international, open-label, randomised, controlled phase 3 trial. *Lancet Oncol* 2018; **19:** 1192–04.

20 Flinn IW, Hillmen P, Montillo M, et al. The phase 3 DUO trial: duvelisib vs ofatumumab in relapsed and refractory CLL/SLL. *Blood* 2018; **132:** 2446–55.

21 Wu YL, Cheng Y, Zhou X, et al. Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): a randomised, open-label, phase 3 trial. *Lancet Oncol* 2017; **18:** 1454–66.

22 Dimopoulos MA, Oriol A, Nahi H, et al. Daratumumab, lenalidomide, and dexamethasone for multiple myeloma. *N Engl J Med* 2016; **375:** 1319–31.

23 Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine J* 2015; **15:** 2188–97.

24 Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The fragility index in multicenter randomized controlled critical care trials. *Crit Care Med* 2016; **44:** 1278–84.

25 Docherty KF, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. How robust are clinical trials in heart failure? *Eur Heart J* 2017; **38:** 338–45.

26 Chase Kruse B, Matt Vassar B. Unbreakable? An analysis of the fragility of randomized trials that support diabetes treatment guidelines. *Diabetes Res Clin Pract* 2017; **134:** 91–105.

27 Edwards E, Wayant C, Besas J, Chronister J, Vassar M. How fragile are clinical trial outcomes that support the CHEST clinical practice guidelines for VTE? *Chest* 2018; **154:** 512–20.

28 Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg* 2019; **154:** 74–79.

29 Tannock IF, Amir E, Booth CM, et al. Relevance of randomised controlled trials in oncology. *Lancet Oncol* 2016; **17:** e560–67.

30 Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat* 2016; **70:** 129–33.

31 Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. *Crit Care* 2004; **8:** 389–94.

32 Carter RE, McKie PM, Storlie CB. The fragility index: a p-value in sheep's clothing? *Eur Heart J* 2017; **38:** 346–48.

33 Booth CM, Del Paggio JC. Approvals in 2016: questioning the clinical benefit of anticancer therapies. *Nat Rev Clin Oncol* 2017; **14:** 135–36.

34 Tibau A, Molto C, Ocana A, et al. Magnitude of clinical benefit of cancer drugs approved by the US Food and Drug Administration. *J Natl Cancer Inst* 2018; **110:** 486–92.

www.thelancet.com/oncology **Published online July 8, 2019 http://dx.doi.org/10.1016/S1470-2045(19)30338-9**