Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

Appendix A1. Alternative endpoints. Endpoints in bold are those evaluated in the case study illustrative example.

	Endpoint	**Definition**	**Studies evaluating these endpoints**
Categorical	Response Evaluation Criteria in Solid Tumors (RECIST) dichotomous response	Complete Response/CR or Partial Response/PR vs. Stable Disease/SD or Progression/PD	Current choice of phase II endpoints in most studies
	Disease Control Rate, DCR	Progression vs. No Progression (CR/PR/Stable), using RECIST cutpoints	Lara et al., *JCO* 2008 Mandrekar et al., *JCO* 2014
	Trichotomous Tumor Response, TriTR	CR/PR vs. Stable vs. Progression, using RECIST cutpoints	Sargent et al. *ASCO Proc* 2008 An, Mandrekar et al., *CCR 2011* Mandrekar et al., *JCO* 2014
Continuous	Log change in tumor size from baseline to time t	log [M(t) - M(0)]	Karrison et al., *JNCI* 2007
	Absolute change, between time points t_1 and t_2	$M(t_2) - M(t_1) / (t_2 – t_1)$	An et al., *JNCI* 2015
	Relative change, between baseline and time t	[M(t) - M(0)] / M(0)	Suzuki et al., *Ann Onc* 2012
	Percent change, between time points t_1 and t_2	$10*[M(t_2) - M(t_1)] / [(t_2 - t_1) * M(t_1)]$	An et al., *JNCI* 2015

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

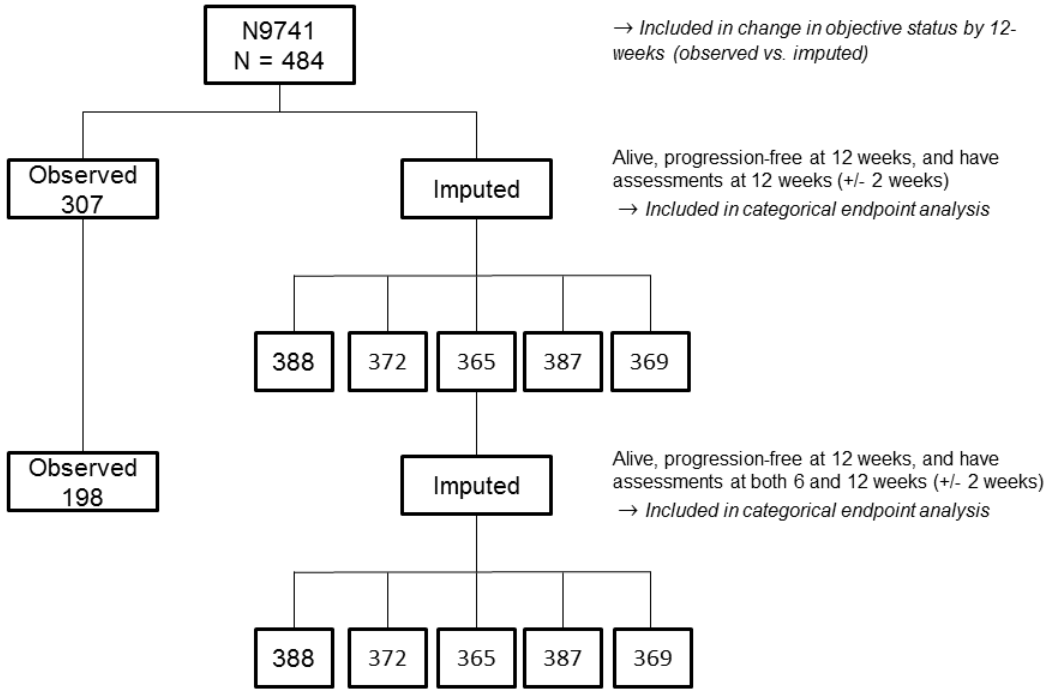
Appendix A2. Calculation of objective status in current work. Measurable lesions up to a maximum number of five at baseline are identified as target lesions. N9741, N9841 and N0026 used criteria other than RECIST 1.1 and possibly have more than five target lesions recorded at baseline. In that scenario, we treated the largest five measurable lesions as target lesions and the rest as non-target lesions. The objective status was calculated only for target lesions (hence measurable) based on RECIST 1.1. Non-target lesions were not considered as they did not have any numerical measurements associated with them.

Objective status	**Criteria**
Complete Response (CR)	Disappearance of all target lesions.
Partial Response (PR)	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters.
Progressive Disease (PD)	At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study. Besides that, the sum must also demonstrate an absolute increase of at least 5 mm. The appearance of new lesions is also considered as progression.
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

Appendix A3. CONSORT Diagram for observed and imputed data for categorical and continuous endpoint analysis, by trial.

Appendix A3, Figure 1. CONSORT Diagram for observed and imputed data for trial N9741 for categorical and continuous endpoint analysis[1].



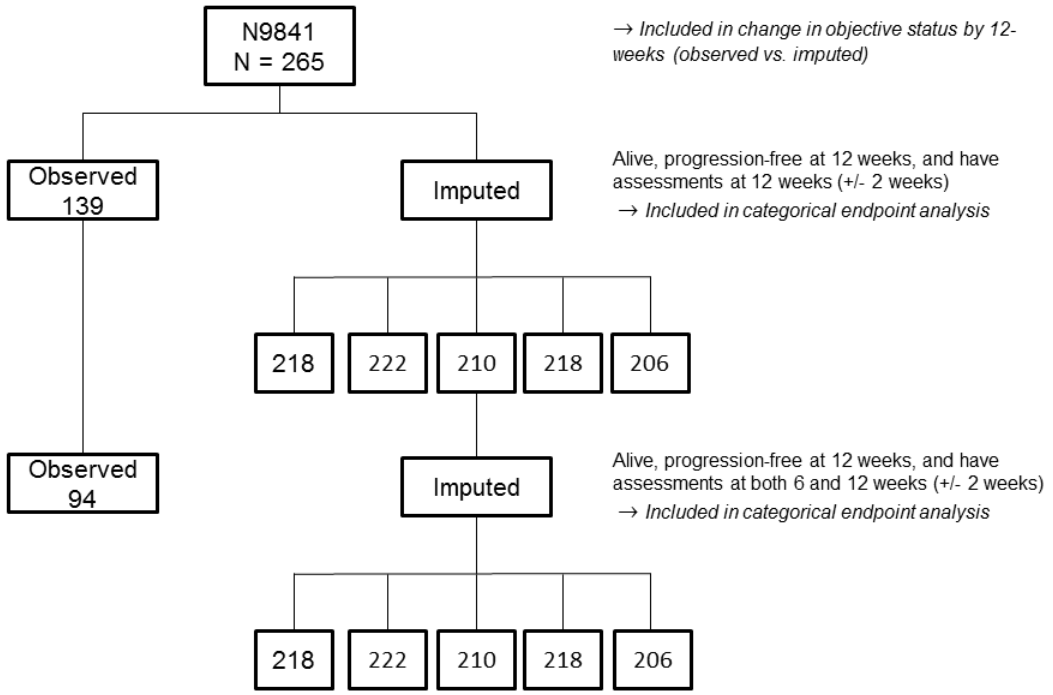
[1]The initial dataset (N=484) includes all patients with at least one baseline assessment **and** one post-baseline assessment after 12 weeks. For the categorical endpoint analysis, the observed dataset (N=307) includes all patients from the initial dataset who **additionally** had no progression at 6-weeks **and** one assessment at 12-weeks. The imputed datasets (N=388, 372, 365, 387, and 369) include all patients who, after imputation of any missing values (observed measurements retained), had at least one baseline assessment **and** no progression at 6-weeks (based on objective status after imputation) **and** one assessment at 12-weeks (necessarily true for all patients, after imputation). See Appendix A4 for a schematic of a dataset with hypothetical patients.

Similarly, for the continuous endpoint analysis, the observed dataset (N=198) includes all patients from the observed dataset for categorical endpoint analysis who **additionally** had one assessment at 6-weeks. The imputed datasets for the

continuous endpoint analysis are identical to those for the categorical endpoint analysis because all patients necessarily have assessments at 6- and 12-weeks with no progression at 6-weeks.

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

Appendix A3, Figure 2. CONSORT Diagram for observed and imputed data for trial N9841 for categorical and continuous endpoint analysis[1].

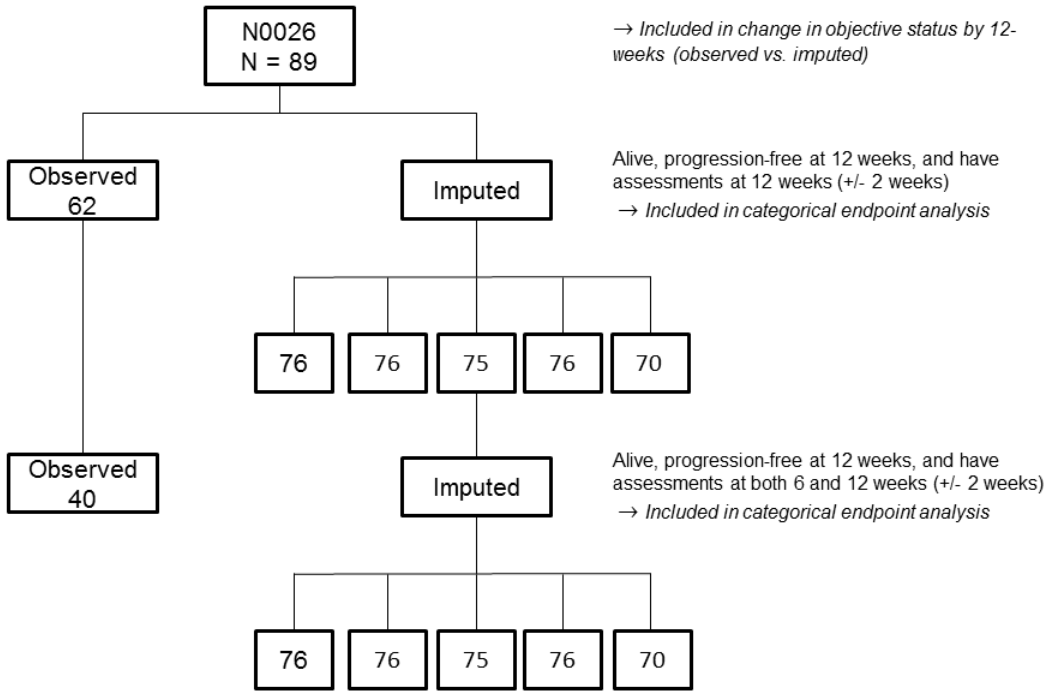


[1]The initial dataset (N=265) includes all patients with at least one baseline assessment **and** one post-baseline assessment after 12 weeks. For the categorical endpoint analysis, the observed dataset (N=139) includes all patients from the initial dataset who **additionally** had no progression at 6-weeks **and** one assessment at 12-weeks. The imputed datasets (N=218, 222, 210, 218, and 206) include all patients who, after imputation of any missing values (observed measurements retained), had at least one baseline assessment **and** no progression at 6-weeks (based on objective status after imputation) **and** one assessment at 12-weeks (necessarily true for all patients, after imputation). See Appendix A4 for a schematic of a dataset with hypothetical patients.

Similarly, for the continuous endpoint analysis, the observed dataset (N=94) includes all patients from the observed dataset for categorical endpoint analysis who **additionally** had one assessment at 6-weeks. The imputed datasets for the continuous endpoint analysis are identical to those for the categorical endpoint analysis because all patients necessarily have assessments at 6- and 12-weeks with no progression at 6-weeks.

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

Appendix A3, Figure 3. CONSORT Diagram for observed and imputed data for trial N0026 for categorical and continuous endpoint analysis[1].



[1]The initial dataset (N=89) includes all patients with at least one baseline assessment **and** one post-baseline assessment after 12 weeks. For the categorical endpoint analysis, the observed dataset (N=62) includes all patients from the initial dataset who **additionally** had no progression at 6-weeks **and** one assessment at 12-weeks. The imputed datasets (N=76, 76, 75, 76, 70) include all patients who, after imputation of any missing values (observed measurements retained), had at least one baseline assessment **and** no progression at 6-weeks (based on objective status after imputation) **and** one assessment at 12-weeks (necessarily true for all patients, after imputation). See Appendix A4 for a schematic of a dataset with hypothetical patients.

Similarly, for the continuous endpoint analysis, the observed dataset (N=40) includes all patients from the observed dataset for categorical endpoint analysis who **additionally** had one assessment at 6-weeks. The imputed datasets for the continuous endpoint analysis are identical to those for the categorical endpoint analysis because all patients necessarily have assessments at 6- and 12-weeks with no progression at 6-weeks.

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

Appendix A4 Detailed explanation of Appendix A2, Figure 1 (*CONSORT Diagram for observed and imputed data for trial N9741 for categorical and continuous endpoint analysis).* Similar explanations apply for Appendix A2, Figures 2 and 3. **Tables below explanation show hypothetical scenarios for observed dataset and an imputed dataset.**

N9741, N = 484

Included if:

- At least one baseline assessment, AND
- One assessment after 10 weeks

N9741 Observed for categorical endpoint analysis, N = 307

Included if:

- At least one baseline assessment AND one assessment after 10 weeks (as above), PLUS
- one assessment at 12 weeks, AND
- no PD at 6-weeks

N9741 Imputed #1 for categorical endpoint analysis, N = 388

Note: All observed measurements retained; only imputed values added

The N=388 patients include:

- Some (though not all) of the N=307 patients in the observed dataset (with no imputation at all if they had complete data, or with some values imputed if any missing data), *e.g. Patient 2 in tables below – note that after imputation, there are 2 possibilities for this type of patient (Patient 2a vs. Patient 2b) which determine whether they're included in the imputed*
- New patients who were not in the observed dataset (i.e. who previously had no 12-w assessment), *e.g. Patient 3 in tables below*

Included if:

- At least one baseline assessment, AND
- No PD at 6-weeks (based on objective status after any imputation), AND
- One assessment at 12 weeks (necessarily true for all patients because of imputation)

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

N9741 Observed for continuous endpoint analysis, N = 198

Included if:

- Meet all criteria for categorical endpoint analysis, AND
- One assessment at 6-weeks

N9741 Imputed #1 for continuous endpoint analysis, N=388
Same as imputed dataset for categorical endpoint analysis because all patients, after imputation, necessarily have assessments a 6- and 12-weeks and no PD at 6-weeks

Observed (missing values denoted by NA)

ID	Time	Lesion 1	Lesion 2	Objective Status, **without** imputation	Included in *observed*?
1	0	X	X		
	6	X	NA	PD	No, 6-w PD
	12	NA	NA		
2	0	X			
	6	NA			
	12	X			Yes, 6-w non-PD and 12-w assessment available
3	0	X			
	6	NA			
	12	NA			No, no 12-w assessment
	Any time t>10w	X			

Imputed (observed values retained; imputed values added and denoted by X*)

ID	Time	Lesion 1	Lesion 2	Objective Status, **with** imputation	Included in *imputed*?
1	0	X	X		
	6	X	X*	PD	No, 6-w PD
	12	X*	X*		
2a	0	X			
	6	X*		PD	
	12	X			No, 6-w PD
2b	0	X			
	6	X*		No PD	Yes, 6-w non-PD
	12	X			
3a	0	X			
	6	X*		PD	No, 6-w PD
	12	X*			
	Any time t>10w	X			
3b	0	X			
	6	X*		No PD	Yes, 6-w non-PD
	12	X*			
	Any time t>10w	X			

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials*, An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

A5. Imputation model.

Tumor size model

By applying the mixed exponential-decay and linear-growth model in Wang et al. (2009) on each individual lesion, we defined the tumor size model as follows:

$$\begin{aligned}\mathrm{TSO}_{ij}(t) &= \mathrm{TS}_i(t)\cdot\exp(\epsilon_{ij})\\ &= \left(\mathrm{BASE}_i\cdot e^{-\mathrm{SR}_i\cdot t}+\mathrm{PR}_i\cdot t\right)\cdot\exp(\epsilon_{ij}),\end{aligned}$$

where

$$\mathrm{BASE}_i=\mathrm{M\_BASE}\cdot\exp(\alpha_i),\quad \mathrm{SR}_i=\mathrm{M\_SR}\cdot\exp(\beta_i),\quad \mathrm{PR}_i=\mathrm{M\_PR}\cdot\exp(\gamma_i),$$

and

$$\alpha_i\overset{iid}{\sim}N(0,\omega^2_{\mathrm{BASE}}),\quad \beta_i\overset{iid}{\sim}N(0,\omega^2_{\mathrm{SR}}),\quad \gamma_i\overset{iid}{\sim}N(0,\omega^2_{PR}),\quad \epsilon_{ij}\overset{iid}{\sim}N(0,\sigma^2_{TS}).$$

Here, $\mathrm{TSO}_{ij}(t)$ represents the observed tumor size at time t for the i^{th} patient and j^{th} lesion. We assumed that lesion measurements are i.i.d. conditional on patient level random effects at any fixed time point. BASE_i, SR_i and PR_i are individual specific baseline tumor size, exponential tumor shrinkage rate and linear tumor progression rate, while M_BASE, M_SR and M_PR are population level average size/rates, respectively.

Equivalently, the model stated above can be re-written as:

$$\begin{aligned}\log(\mathrm{TSO}_{ij}(t)) &= \log\left(\mathrm{BASE}_i\cdot e^{-\mathrm{SR}_i\cdot t}+\mathrm{PR}_i\cdot t\right)+\epsilon_{ij}\\ &= \log\left(\exp(\log(\mathrm{BASE}_i))\cdot e^{-\exp(\log(\mathrm{SR}_i))t}+\exp(\log(\mathrm{PR}_i))\cdot t\right)+\epsilon_{ij}.\end{aligned}\tag{1}$$

In terms of model estimation, we adopted the Bayesian approach by using the `brms` package in R. Let

$$\begin{aligned}\mathrm{LTSO}_{ij}(t)&\equiv\log(\mathrm{TSO}_{ij}(t)),\\ \mathrm{LBASE}_i&\equiv\log(\mathrm{BASE}_i)=\log(\mathrm{M\_BASE})+\alpha_i,\\ \mathrm{LSR}_i&\equiv\log(\mathrm{SR}_i)=\log(\mathrm{M\_SR})+\beta_i,\\ \mathrm{LPR}_i&\equiv\log(\mathrm{PR}_i)=\log(\mathrm{M\_PR})+\gamma_i.\end{aligned}$$

Plugging these terms back in Eq (1) would give

$$\mathrm{LTSO}_{ij}(t)=\log\left(\exp(\mathrm{LBASE}_i)\cdot e^{-\exp(\mathrm{LSR}_i)\cdot t}+\exp(\mathrm{LPR}_i)\cdot t\right)+\epsilon_{ij},\tag{2}$$

where

$$\begin{aligned}\mathrm{LBASE}_i&\overset{iid}{\sim}N(\log(\mathrm{M\_BASE}),\omega^2_{\mathrm{BASE}}),\\ \mathrm{LSR}_i&\overset{iid}{\sim}N(\log(\mathrm{M\_SR}),\omega^2_{\mathrm{SR}}),\\ \mathrm{LPR}_i&\overset{iid}{\sim}N(\log(\mathrm{M\_PR}),\omega^2_{\mathrm{PR}}).\end{aligned}$$

Moreover, we added a small positive disturbance whenever the observed tumor measurement is zero in order for the log transformation in Eq (2) to work. The estimates of parameters and their estimation precisions were summarized in Table 1.

The estimated baseline population average tumor size were similar across studies. The population level effect, i.e. log(M_BASE), ranged from 1.20-1.35, and the individual level effect, i.e. ω_{BASE}, ranged from 0.24-0.29. Study N0026 had the smallest population level shrinkage and progression rates, while the largest interpatient variability. Meanwhile, the estimation precisions of study N0026 were considerably worse than the other two, which might due to the smaller sample size of the study.

Table 1: Parameter estimates and their estimation precisions(SE) by `brms`.

Study	log(M_BASE)	log(M_SR)	log(M_PR)	
N9741	1.35 (0.02)	-6.46 (0.19)	-9.26 (0.30)	
N9841	1.30 (0.03)	-7.47 (0.47)	-8.69 (0.43)	
N0026	1.20 (0.04)	-9.86 (2.25)	-11.27 (2.48)	
Study	ω_{BASE}	ω_{SR}	ω_{PR}	σ_{TS}
N9741	0.24 (0.03)	1.90 (0.15)	2.02 (0.19)	1.06 (0.01)
N9841	0.28 (0.04)	2.38 (0.33)	1.89 (0.28)	1.00 (0.01)
N0026	0.29 (0.04)	3.34 (1.35)	2.54 (1.27)	0.63 (0.02)

Multiple imputation

The imputation steps we adopted in the paper to create each imputed data set are summarized below:

- Step 1: Obtain parameter estimates using `brms` package including all the available tumor measurements before data realignment, for 577, 377, and 126 patients from N9741, N9841, and N0026, respectively.
- Step 2: For patient i with $m(m > 0)$ missing tumor measurements at 6-weeks and/or 12-weeks, make random draws for patient specific random effects, i.e. LBASE_i, LSR_i and LPR_i with plug-in estimates obtained from Step 1.
- Step 3: For each of m missing tumor measurements, make a random draw from the error distribution $N(0, \hat{\sigma}^2_{TS})$.
- Step 4: Assemble simulated patient specific random effects from Step 2 and noise(s) from Step 3 according to Eq (1) to complete the tumor measurements for patient i.
- Step 5: Repeat Step 2 - 5 for all patients with incomplete data.

In addition to the simulation-based imputation approach illustrated above, we also tried another prediction-based approach. After fitting the mixed exponential-decay and linear-growth model, we imputed each missing tumor measurement with its predicted value from the posterior predictive distribution using the `predict.brmsfit` function. For illustration purpose, we displayed the OS predictive ability metrics for a single imputed data set based on this prediction-based approach in Table 2.

Table 2: OS predictive ability metrics for a single imputed data set based on prediction-based imputation approach, by study and by endpoint.

Study	Endpoint	Events/Total	C-Index	95% CI	HL Stat	LRT P-val	BIC
N9741	Dichotomous	426/442	0.5902	(0.5588-0.6216)	0.0370	< 0.0001	4424.01
	Trichotomous	426/442	0.5896	(0.5582-0.6210)	0.0395	< 0.0001	4430.06
	DCR	426/442	0.5802	(0.5488-0.6116)	0.0295	0.0014	4435.79
	Absolute	426/442	0.5921	(0.5607-0.6235)	0.0316	0.0001	4441.99
	Relative	426/442	0.6019	(0.5705-0.6333)	0.0759	< 0.0001	4434.74
N9841	Dichotomous	228/242	0.5956	(0.5529-0.6383)	0.1006	0.0017	2106.05
	Trichotomous	228/242	0.5905	(0.5478-0.6332)	0.1070	0.0029	2110.26
	DCR	228/242	0.5968	(0.5541-0.6395)	0.0733	0.0011	2105.20
	Absolute	228/242	0.6239	(0.5812-0.6666)	0.0661	< 0.0001	2101.50
	Relative	228/242	0.6147	(0.5720 -0.6574)	0.0815	0.0002	2111.07
N0026	Dichotomous	79/86	0.5345	(0.4622-0.6068)	0.2649	0.7947	589.71
	Trichotomous	79/86	0.5342	(0.4619-0.6065)	0.1972	0.6928	593.08
	DCR	79/86	0.5502	(0.4779-0.6225)	0.3432	0.5181	588.85
	Absolute	79/86	0.6168	(0.5443-0.6893)	0.2648	0.1724	595.56
	Relative	79/86	0.5749	(0.5024-0.6474)	0.2085	0.8658	601.40

Appendix for *Missing Tumor Measurement (TM) Data in the Search for Alternative TM-based Endpoints in Cancer Clinical Trials,* An MW, Tang J, Grothey A, Sargent DJ, Ou FS, Mandrekar SJ

Appendix A6. Statistical measures of predictive ability.

Measure	**Definition**	**Description**	**Reference**
Concordance (c-) Index	Proportion of all pairs of patients *i,j* for whom it can be determined that $T_i < T_j$ who are concordant, i.e. predicted hazard for patient *i* < predicted hazard for patient *j*	Measure of model discrimination, i.e. how well the model differentiates between different outcomes	Harrell et al., *JAMA 1982*
Hosmer-Lemeshow-type statistic	Group patients into deciles of their predicted probabilities from a Cox model. Within each decile, calculate average predicted probability ("expected") and the Kaplan–Meier estimate ("observed"). Sum over deciles the squared differences between expected and observed probabilities.	Measure of model calibration, i.e. how well the model-predicted 1-year survival probabilities match observed 1-year survival probabilities	D'Agostino et al., *Handbook of Statistics 2004*
Bayesian Information Criteria (BIC)	-2 log Likelihood + log(n) * k, where n = sample size and k = number of parameters in model	Measure of model fit, penalizing for sample size and number of parameters in model	