How to systematically approach the evaluation of a randomized controlled trial

Dr Indranil Mallick

Tata Medical Center, Kolkata

What is a randomized controlled trial?

A study design that **randomly** assigns participants into an **experimental** group or a **control** group.

As the study is conducted, the only expected difference between the control and experimental groups in a randomized controlled trial (RCT) is the intervention being studied.

You don’t need to be a expert or statistician to interpret an RCT

The Question

Control Arm

Endpoints

Experimental Arm

Is the question correct and clinically meaningful?

The Methods

Patient Selection

Randomization

Sample size

Control Arm Rx

Exp Arm Rx

Are the methods appropriate?

The Results

Analysis and Results

Are the results correctly analyzed?

Discussion

Context

Are the results explainable and merit implementation?

The Question

Control Arm

Endpoints

Experimental Arm

Is the question correct and clinically meaningful?

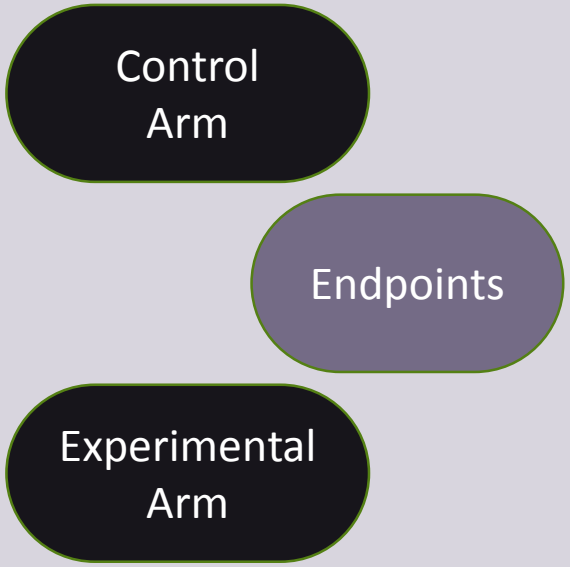
A wrong question cannot have a right answer

Is the control treatment the current standard of care?

Is the experimental treatment logical, safe and implementable?

Are the endpoints clinically meaningful?

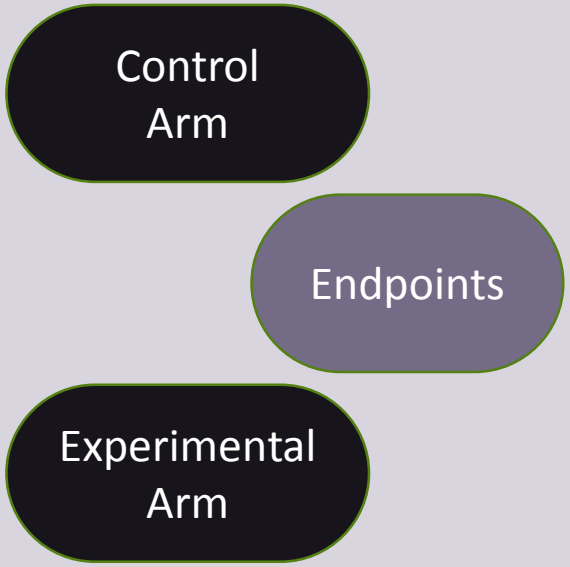
The Question



Is the control treatment the current standard of care?

- Unless the control arm represents the current standard of care, the trial may not provide a clinically meaningful answer.
- Check for details:
  - Drug dose schedules
  - Radiation volumes, dose-fractionation, techniques
  - Surgical details
- The control arm may need updating during the course of the study if standard of care changes.

The Question

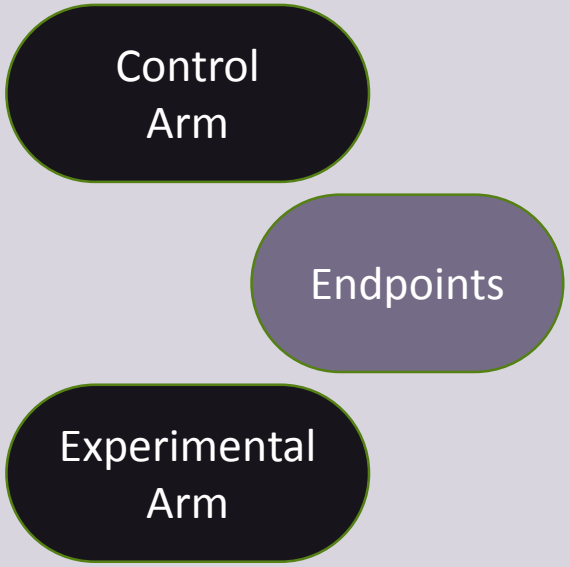


Is the question correct and clinically meaningful?

Is the experimental treatment logical, safe and implementable?

- Is there a biological/clinical justification in using this experimental arm?
- Is there Phase I/II data that suggests safety/efficacy?
- Is the treatment schedule consistent with known usage?
  - Drug dose schedules
  - Radiation volumes, dose-fractionation, techniques
  - Surgical details

The Question



Is the question correct and clinically meaningful?

Are the endpoints valid and clinically meaningful?

- **Clinically meaningful endpoints** – overall survival and quality of life.
- **Surrogate endpoints often do not correlate with OS** – response rates, disease-free survival, progression free survival, biochemical control, metastasis free survival.
- **Toxicity endpoints** – valid only if reported by patients
- **Secondary endpoints** – only hypothesis generating

Long term results of RTOG 91-11

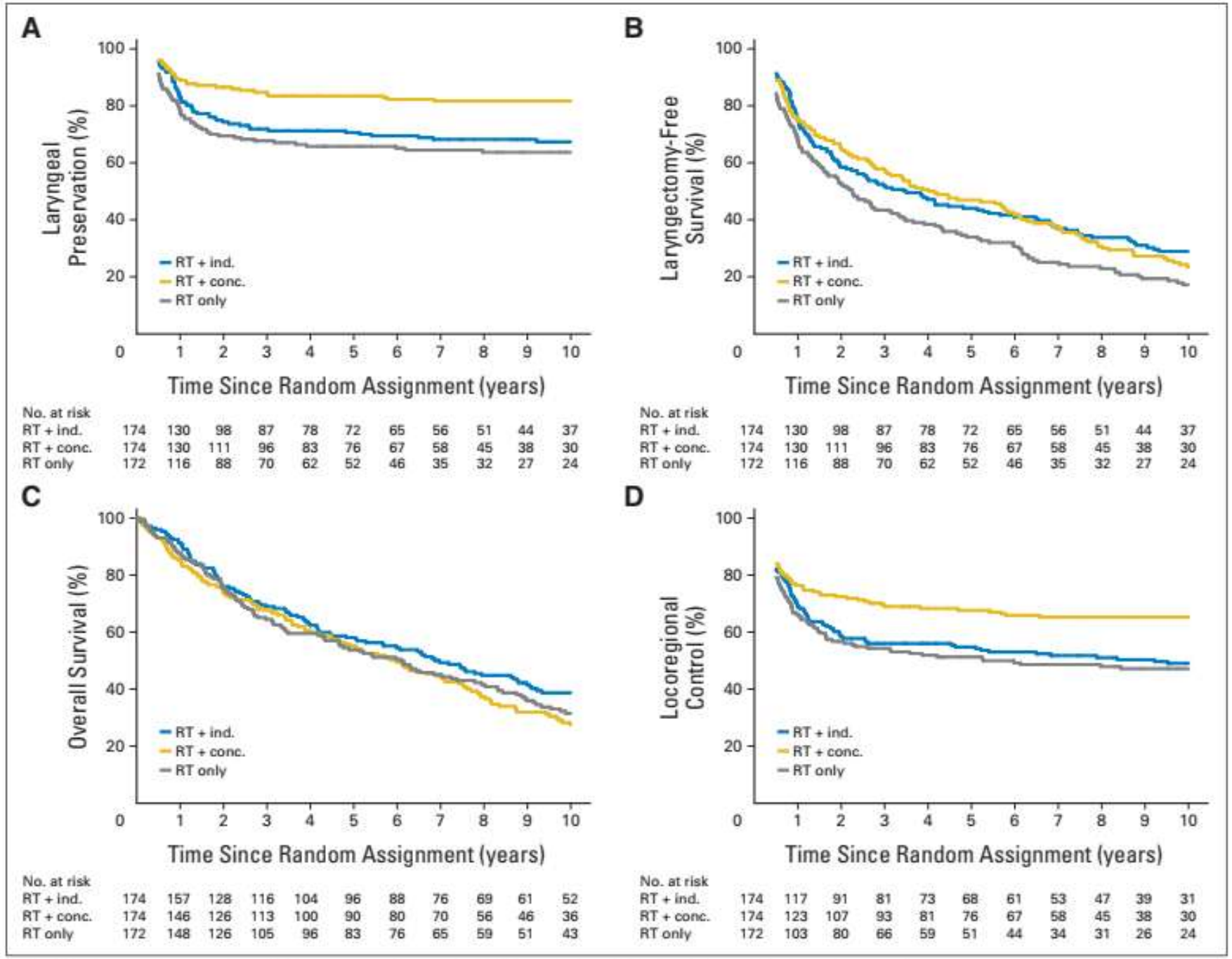


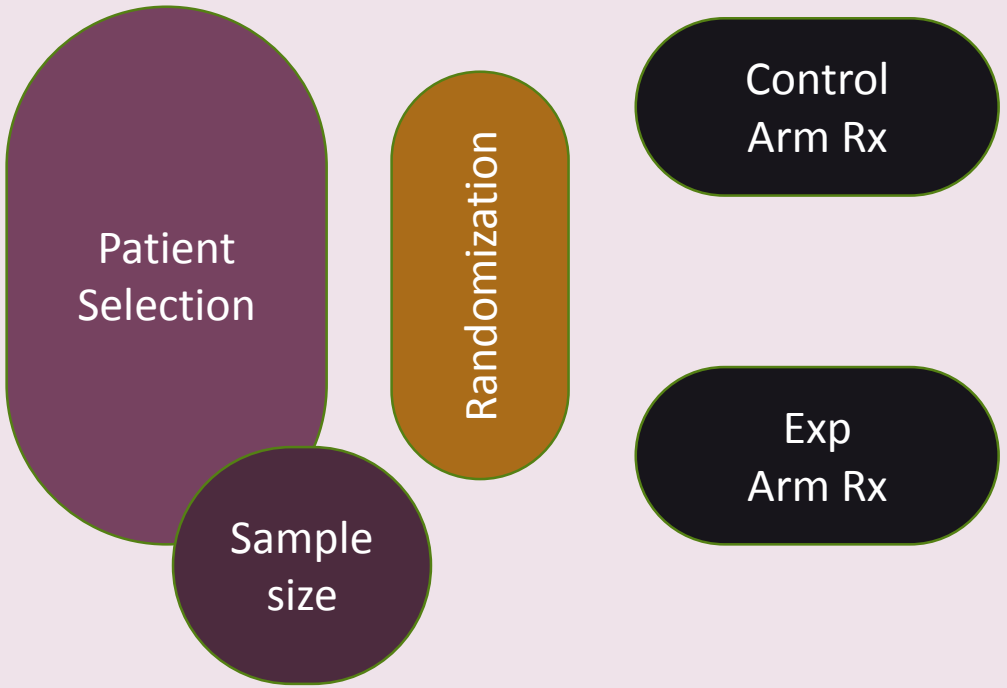
Fig 2. (A) Laryngeal preservation, (B) laryngectomy-free survival, (C) overall survival, and (D) locoregional control according to treatment group. conc., concomitant; ind., induction; RT, radiation therapy.

Forastiere, JCO 2013

The Methods

Is the patient selection criteria externally valid?

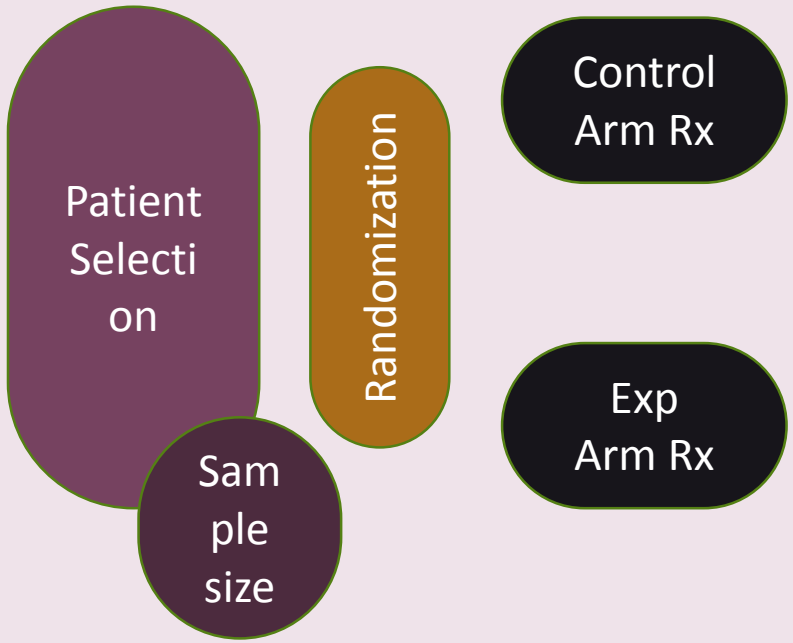
Is the sample size appropriately calculated?



Is the process of randomization robust?

Was the treatment processes as specified and quality assured?

The Methods



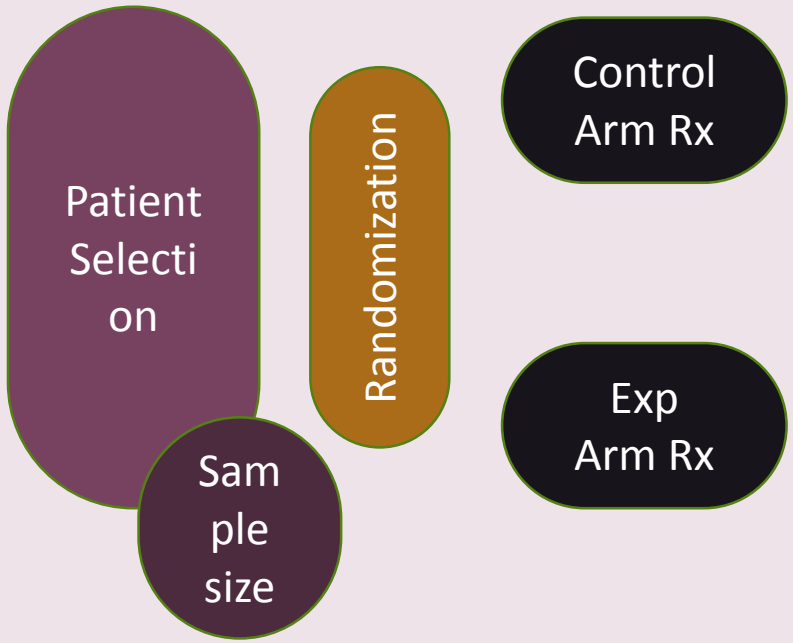
Is the patient selection criteria reflective of common practice?

Inclusion criteria – is it accommodating the range of stages that matter?

Exclusion criteria – is it excluding a lot of patients with comorbidities?

Is this the true population where you are looking to use this new treatment?

The Methods



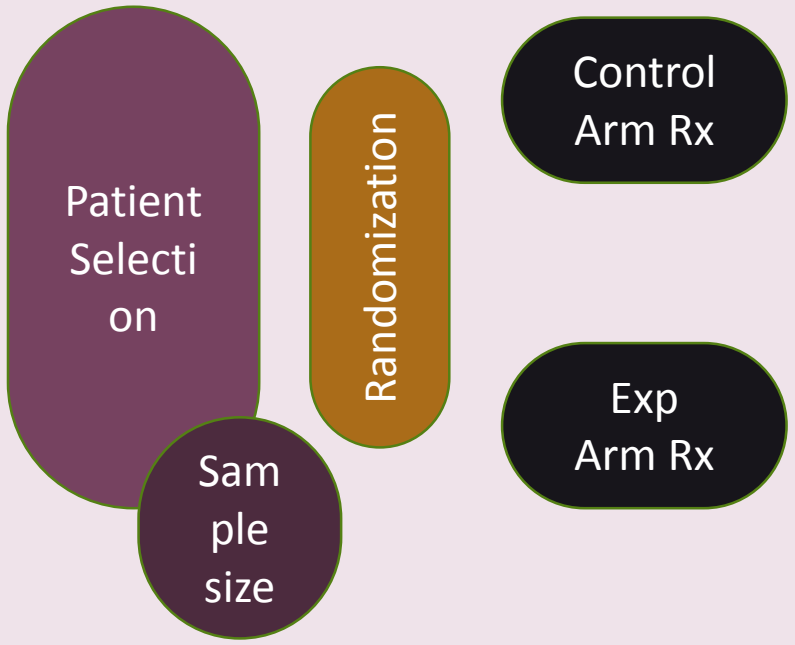
Calculated for the primary endpoint.

Depends upon:

- **The relative difference expected (Hazard Ratio, or likelihood of event in exp vs control arm)**
- **The specified type I and type II errors/power**
- Duration of recruitment and follow up

Is the sample size appropriately calculated?

The Methods

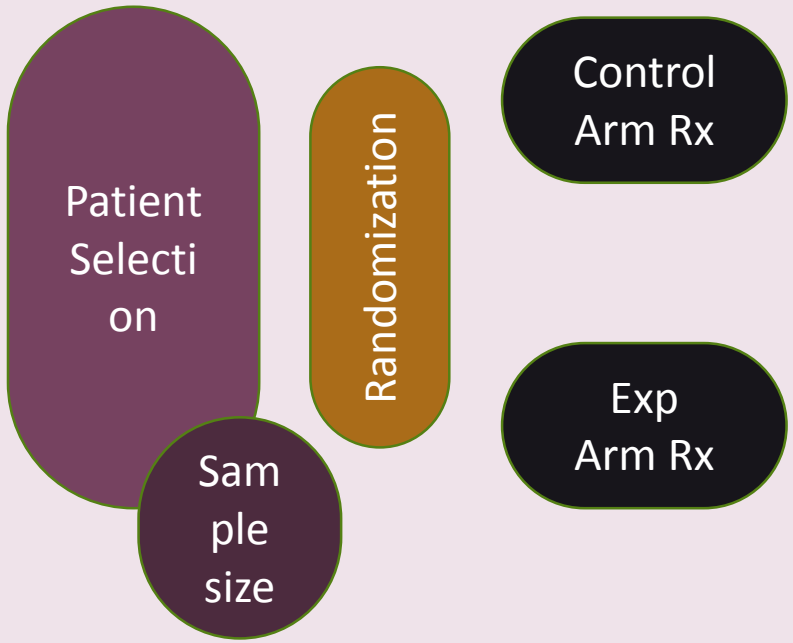


Is the process of randomization +/- blinding robust?

Look for the randomization method, especially in smaller single-institution RCTs. Confirm allocation concealment.
Is the randomization stratified using important variables? Or randomization with minimization?

Blinding reduces biases in reporting, assessment and surveillance

The Methods

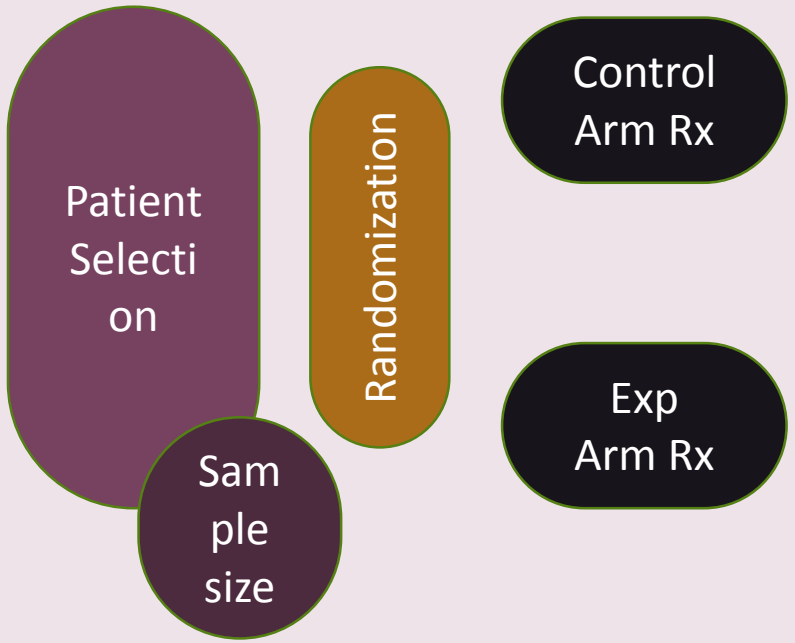


The importance of treatment QA is underestimated especially for multi-institutional studies:

- Pathological/molecular characteristics
- Radiation treatment planning
- Surgical techniques/training
- Drug storage/administration/PD-PK studies

Was the treatment processes as specified and quality assured?

The Methods



What happens to patients in the control arm if they fail?

Are the patients offered the standard salvage therapy (if necessary, with the therapy in the experimental arm – 'crossover')

How soon are they offered salvage therapy?

The Results

What are the patient and disease characteristics ?

Does the analysis accounts for all patients?

Is the treatment compliance and toxicity profile reported?

Is it an intention-to-treat analysis?

Analysis and Results

What are the patient and disease characteristics?

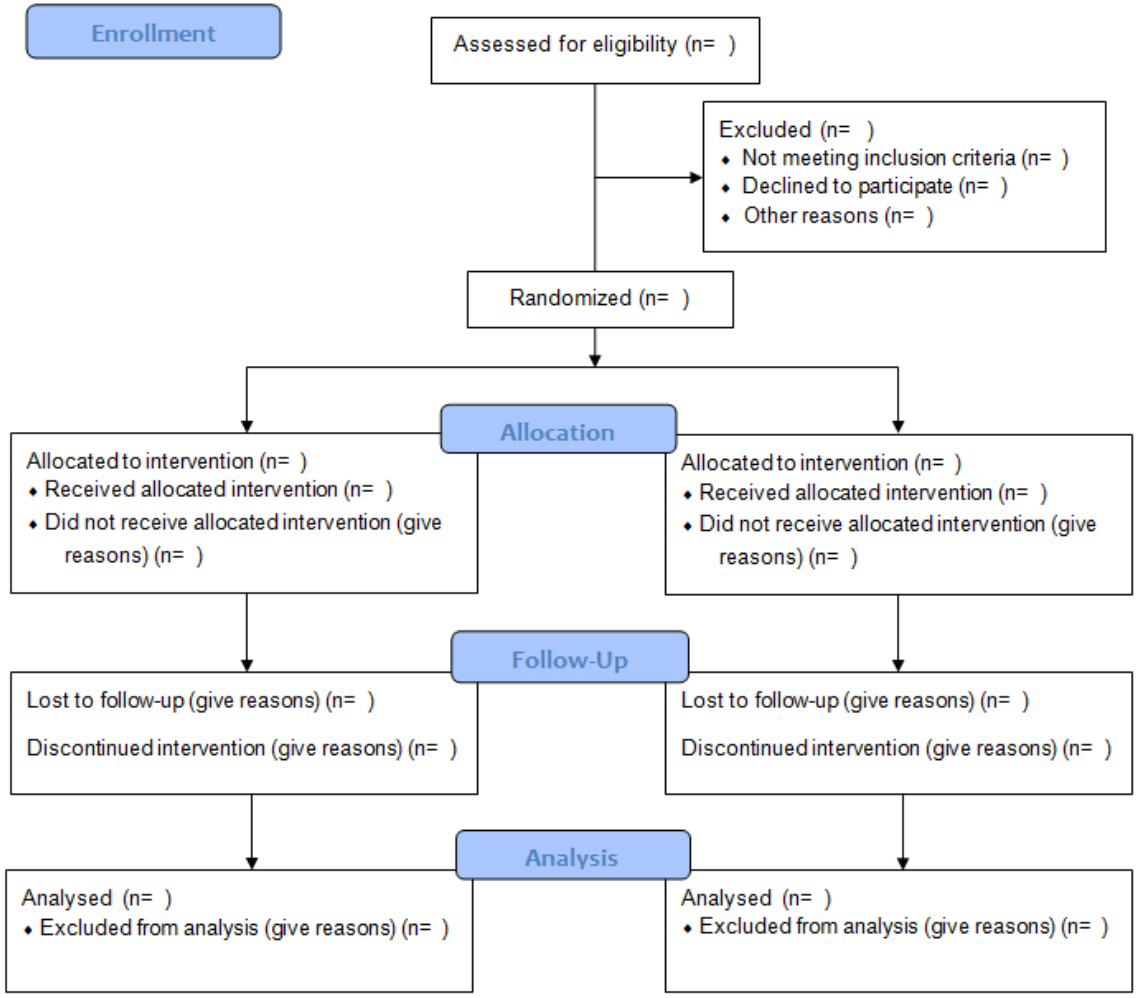
- Is there a balance between the arms in terms of stage and risk factors of recurrence? (if there is no stratification in randomization)
- Do they reflect the usual distribution in your practice?

	Control (n=1029)	Radiotherapy (n=1032)
Age at randomisation (years)	68 (63–73)	68 (63–73)
Range	37–86	45–87
WHO performance status		
0	732 (71%)	734 (71%)
1–2	297 (29%)	298 (29%)
Pain from prostate cancer		
Absent	820 (81%)	844 (83%)
Present	198 (19%)	170 (17%)
Missing data	11	18
Previous notable health issues		
Myocardial infarction	67 (7%)	57 (6%)
Cerebrovascular disease	29 (3%)	30 (3%)
Congestive heart failure	5 (<1%)	8 (1%)
Angina	46 (4%)	51 (5%)
Hypertension	408 (40%)	440 (43%)
Missing data	5	8
T category at randomisation		
T0	0 (0%)	2 (<1%)
T1	12 (1%)	12 (1%)
T2	84 (9%)	89 (9%)
T3	585 (62%)	603 (63%)
T4	260 (28%)	246 (26%)
TX	88	80
N category at randomisation		
N0	345 (36%)	344 (36%)
N+	620 (64%)	620 (64%)
NX	64	68

Analysis and Results

CONSORT
TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram



Does the analysis accounts for all patients?

Important differences in the proportion of patients who are lost to follow up and analyzed – poor quality of a study

Analysis and Results

The Results

Analysis and Results

Table 2. Adverse Events.*

Event	Nivolumab plus Chemotherapy (N=176)		Chemotherapy Alone (N=176)	
	Any Grade	Grade 3 or 4	Any Grade	Grade 3 or 4
Adverse events of any cause — no. (%)†				
All	163 (92.6)	72 (40.9)	171 (97.2)	77 (43.8)
Leading to discontinuation of treatment	18 (10.2)	10 (5.7)	20 (11.4)	7 (4.0)
Serious	30 (17.0)	19 (10.8)	24 (13.6)	17 (9.7)
Treatment-related adverse events — no. (%)†				
All	145 (82.4)	59 (33.5)	156 (88.6)	65 (36.9)
Leading to discontinuation of treatment	18 (10.2)	10 (5.7)	17 (9.7)	6 (3.4)
Serious	21 (11.9)	15 (8.5)	18 (10.2)	14 (8.0)
Death‡	0	—	3 (1.7)	—
Surgery-related adverse events — no./total no. (%)§	62/149 (41.6)	17/149 (11.4)	63/135 (46.7)	20/135 (14.8)

Is the treatment compliance and toxicity profile reported?

Is it an intention-to-treat analysis?

- **ITT analysis -** once randomized, always analysed in the randomized group
- Regardless of their
  - adherence with the entry criteria
  - treatment they actually received
  - subsequent withdrawal from treatment or deviation from the protocol
- Non ITT analyses removes the benefit of the balancing provided by randomization.
  - **Per-protocol analysis**: only those patients who were treated according to protocol
  - **As-treated analysis**: analysed on the basis of the treatment they actually received.

The Results

- **Focus on the primary endpoint**
- **Look at the hazard ratio –** point estimate – likelihood of relative benefit
- **Look at the confidence intervals –** estimate of precision of the point estimate
- **Look at the Kaplan Meier curve** do the curves truly reflect a difference.
- **Look at the p-value last**
- **Look at secondary endpoints** in context

Table 2. Hazard Ratios for Efficacy End Points

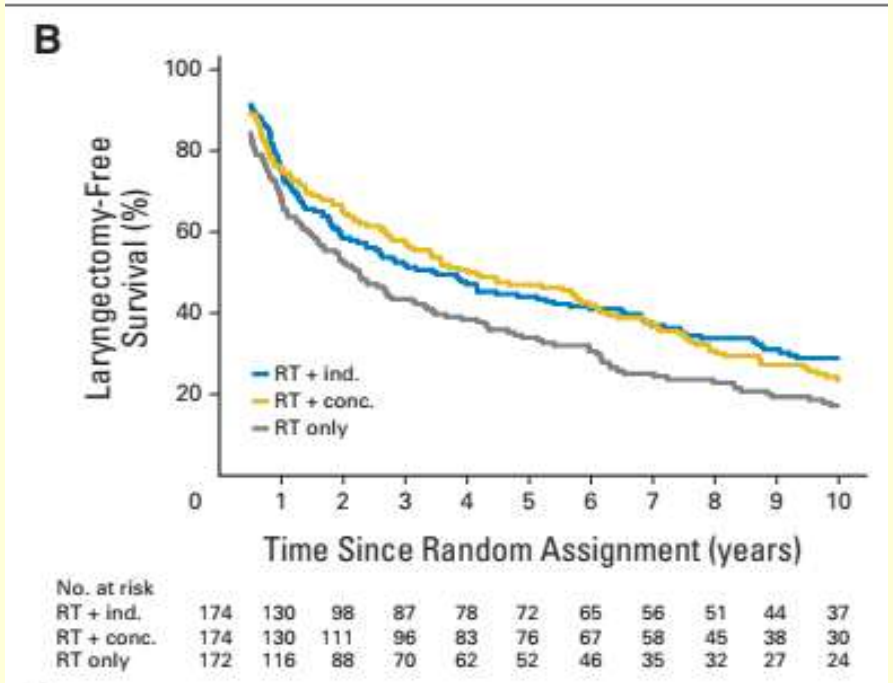
End Point	Hazard Ratio*	95% CI	*P*
Laryngectomy-free survival			
RT + concomitant *v* RT + induction	1.05	0.83 to 1.34	.68
RT alone *v* RT + induction	1.33	1.05 to 1.69	.02
RT + concomitant *v* RT alone	0.78	0.61 to 0.98	.03
Laryngeal preservation			
RT + concomitant *v* RT + induction	0.58	0.37 to 0.89	.005
RT alone *v* RT + induction	1.26	0.88 to 1.82	.35
RT + concomitant *v* RT alone	0.46	0.30 to 0.71	< .001
Local control			
RT + concomitant *v* RT + induction	0.66	0.47 to 0.93	.006
RT alone *v* RT + induction	1.18	0.87 to 1.60	.50
RT + concomitant *v* RT alone	0.57	0.40 to 0.80	< .001
Locoregional control			
RT + concomitant *v* RT + induction	0.66	0.48 to 0.92	.0037
RT alone *v* RT + induction	1.13	0.84 to 1.52	.72
RT + concomitant *v* RT alone	0.59	0.43 to 0.82	.0015
Distant control			
RT + concomitant *v* RT + induction	1.11	0.66 to 1.86	.88
RT alone *v* RT + induction	1.59	0.99 to 2.58	.06
RT + concomitant *v* RT alone	0.69	0.43 to 1.11	.08
Disease-free survival			
RT + concomitant *v* RT + induction	0.98	0.78 to 1.24	.88
RT alone *v* RT + induction	1.26	1.00 to 1.58	.06
RT + concomitant *v* RT alone	0.78	0.62 to 0.98	.04
Overall survival			
RT + concomitant *v* RT + induction	1.25	0.98 to 1.61	.08
RT alone *v* RT + induction	1.15	0.89 to 1.47	.29
RT + concomitant *v* RT alone	1.08	0.85 to 1.39	.53

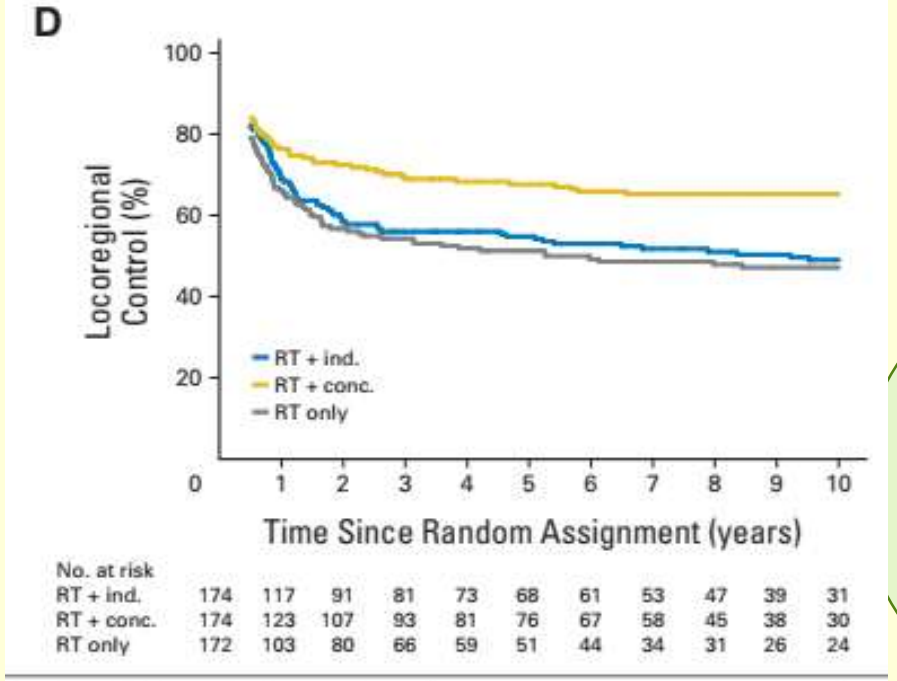
Analysis and Results

Looking at results

The Results

- **Focus on the primary endpoint**
- **Look at the hazard ratio –** point estimate – likelihood of relative benefit
- **Look at the confidence intervals –** estimate of precision of the point estimate
- **Look at the Kaplan Meier curve** do the curves truly reflect a difference.
- **Look at the p-value last**
- **Look at secondary endpoints** in context





vival, and (D) locoregional control according to treatment group. conc., concomitant

Analysis and Results

Looking at results

Discussion

Are differences with other studies explained?

Requires subject matter expertise

Are the right conclusions being drawn?

'Tolerable safety profile'
'Standard of care'

Statistical significance vs. clinical significance

Have the conflicts of interest been reported?

Context

Statistical significance vs. clinical significance

Erlotinib Plus Gemcitabine Compared With Gemcitabine Alone in Patients With Advanced Pancreatic Cancer: A Phase III Trial of the National Cancer Institute of Canada Clinical Trials Group

Malcolm J. Moore, David Goldstein, John Hamm, Arie Figer, Joel R. Hecht, Steven Gallinger, Heather J. Au, Pawel Murawa, David Walde, Robert A. Wolff, Daniel Campos, Robert Lim, Keyue Ding, Gary Clark, Theodora Voskoglou-Nomikos, Mieke Ptasynski, and Wendy Parulekar

Results

A total of 569 patients were randomly assigned. Overall survival based on an intent-to-treat analysis was significantly prolonged on the erlotinib/gemcitabine arm with a hazard ratio (HR) of 0.82 (95% CI, 0.69 to 0.99; $P = .038$, adjusted for stratification factors; median 6.24 months *v* 5.91 months). One-year survival was also greater with erlotinib plus gemcitabine (23% *v* 17%;

Practice points

- Don’t interpret an RCT (or any other clinical study) by the abstract alone
- Critical analysis is a systematic process and an essential skill – read, practice, discuss and argue
- Read editorials and letters to editor in the journals
- Don’t take studies on their face value