

Key points to consider for the statistical analysis of Patient Reported Outcomes

Francesco Cottone, PhD

Italian Group for Adult Hematologic Diseases (GIMEMA),
Data Center and Health Outcomes Research Unit, Rome, Italy.

Patient Reported Outcomes (PROs)

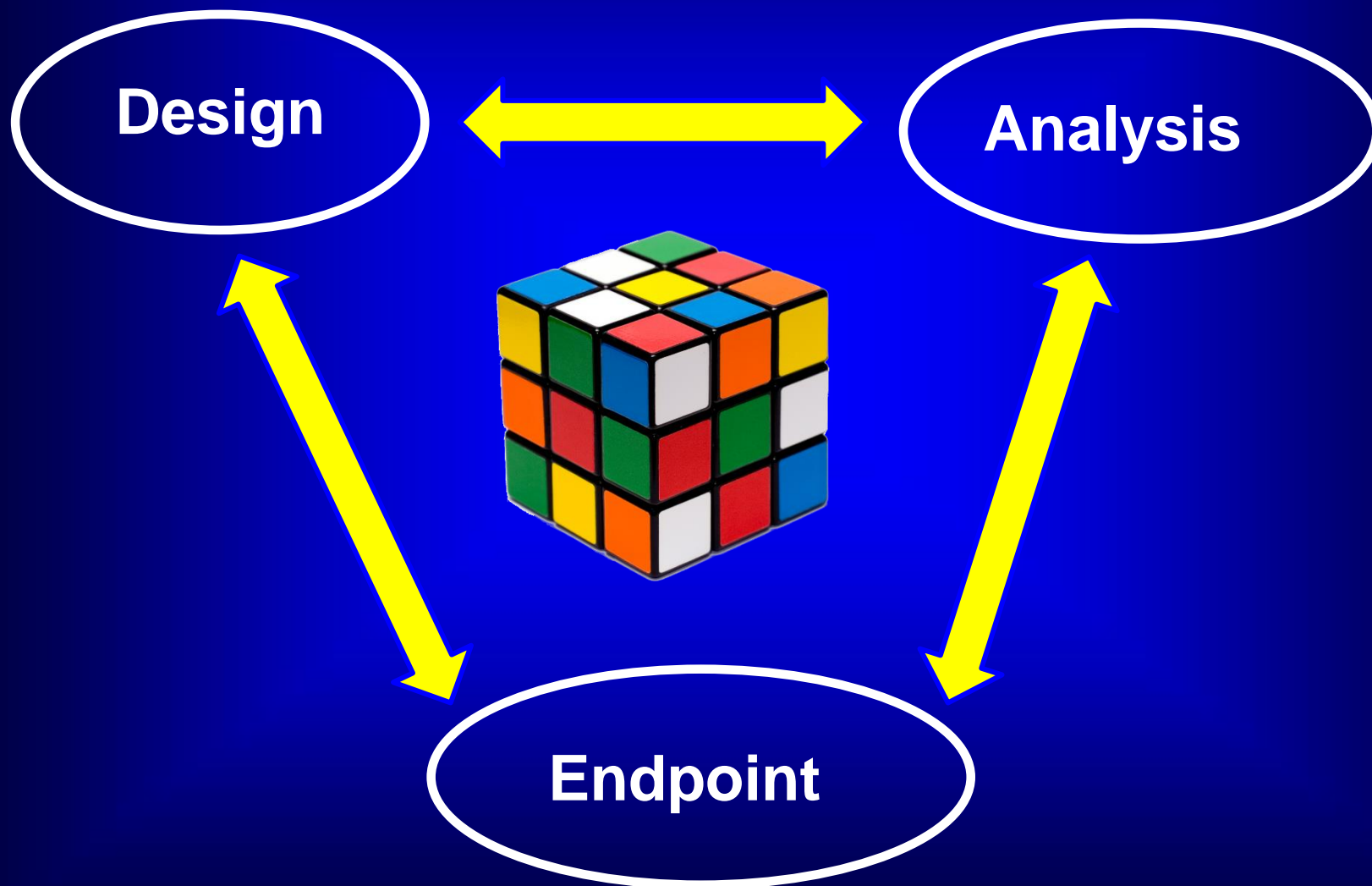
1. A measurement of any aspect of a patient's health status.
2. It comes directly from the patient, without the interpretation of patient's responses by a physician or anyone else.

FDA definition (<http://www.fda.gov/cder/guidance/5460dft.pdf>)

Questionnaire

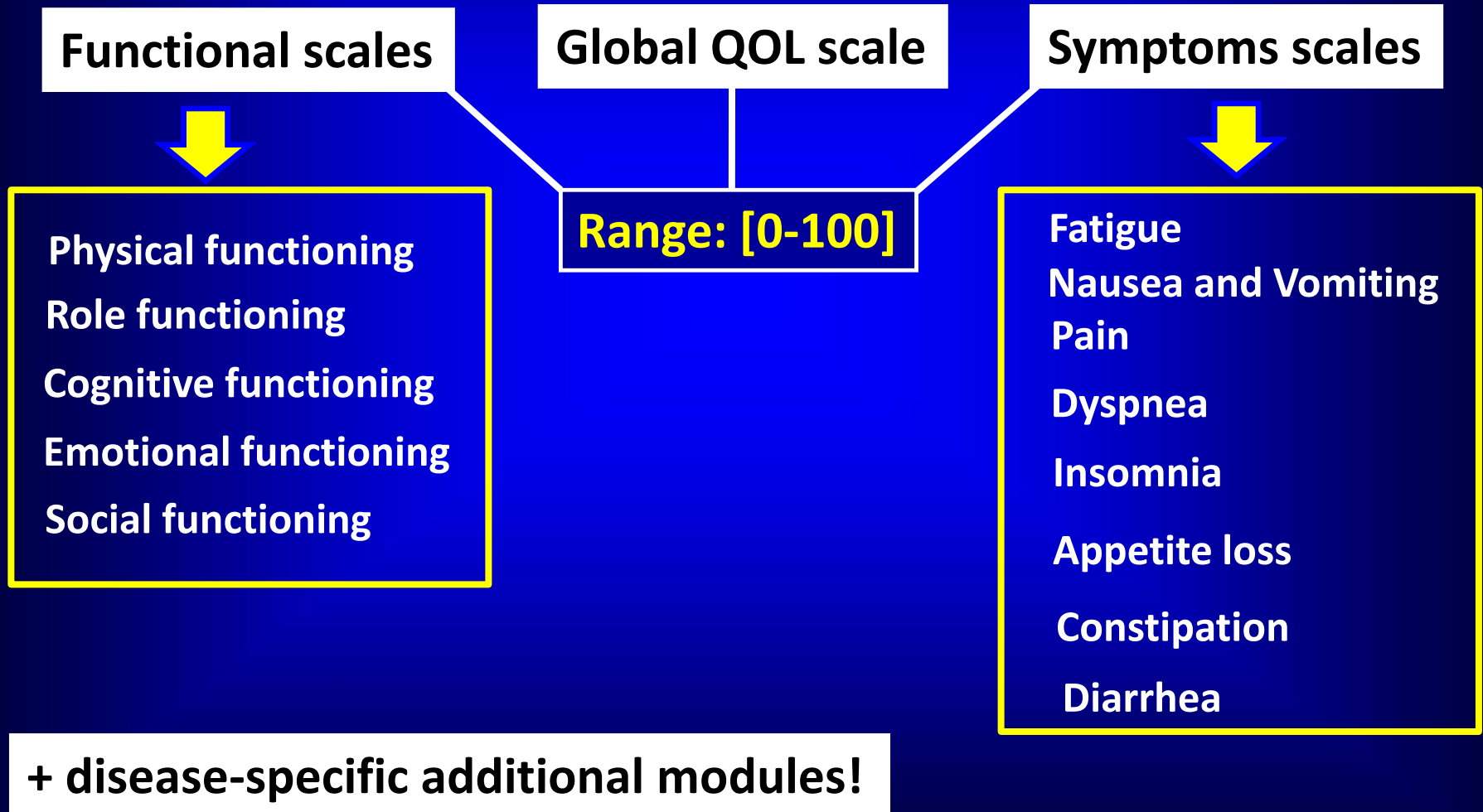
- **Multi-dimensional**
- **Patient Self-rated**
- **Scientifically sound**

Statistical issues

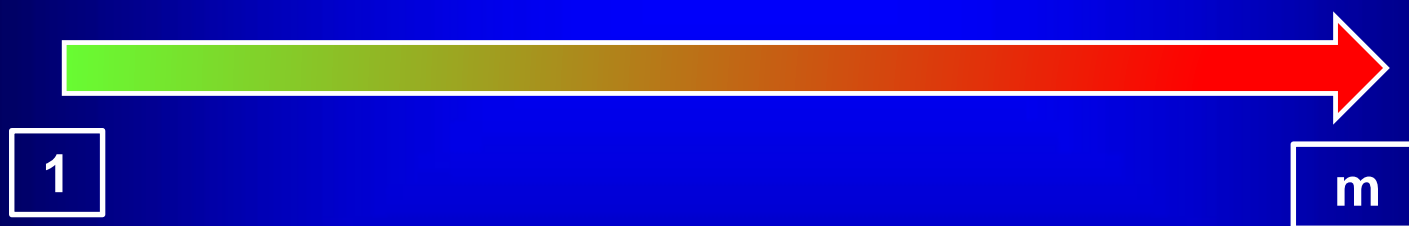
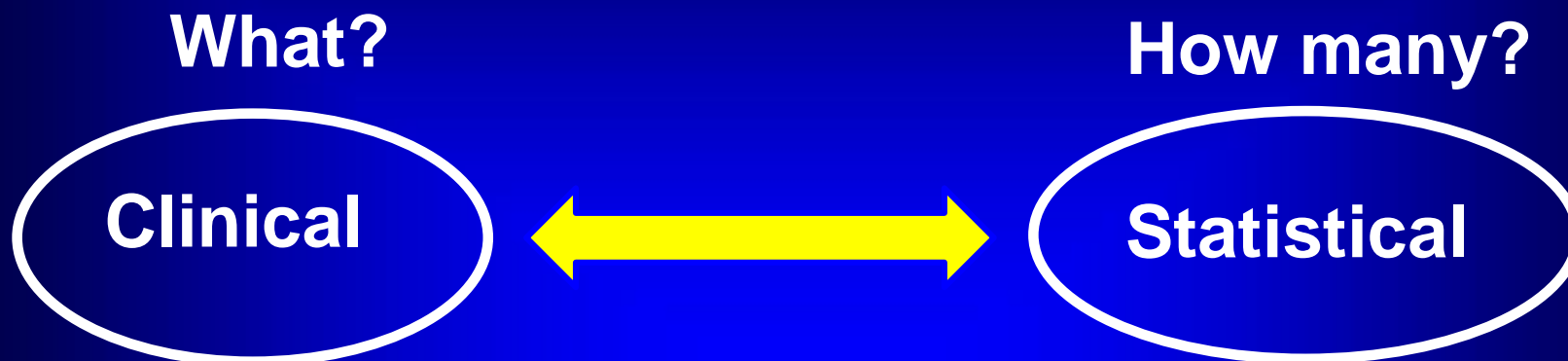


Multiple outcomes

EORTC QLQ-C30

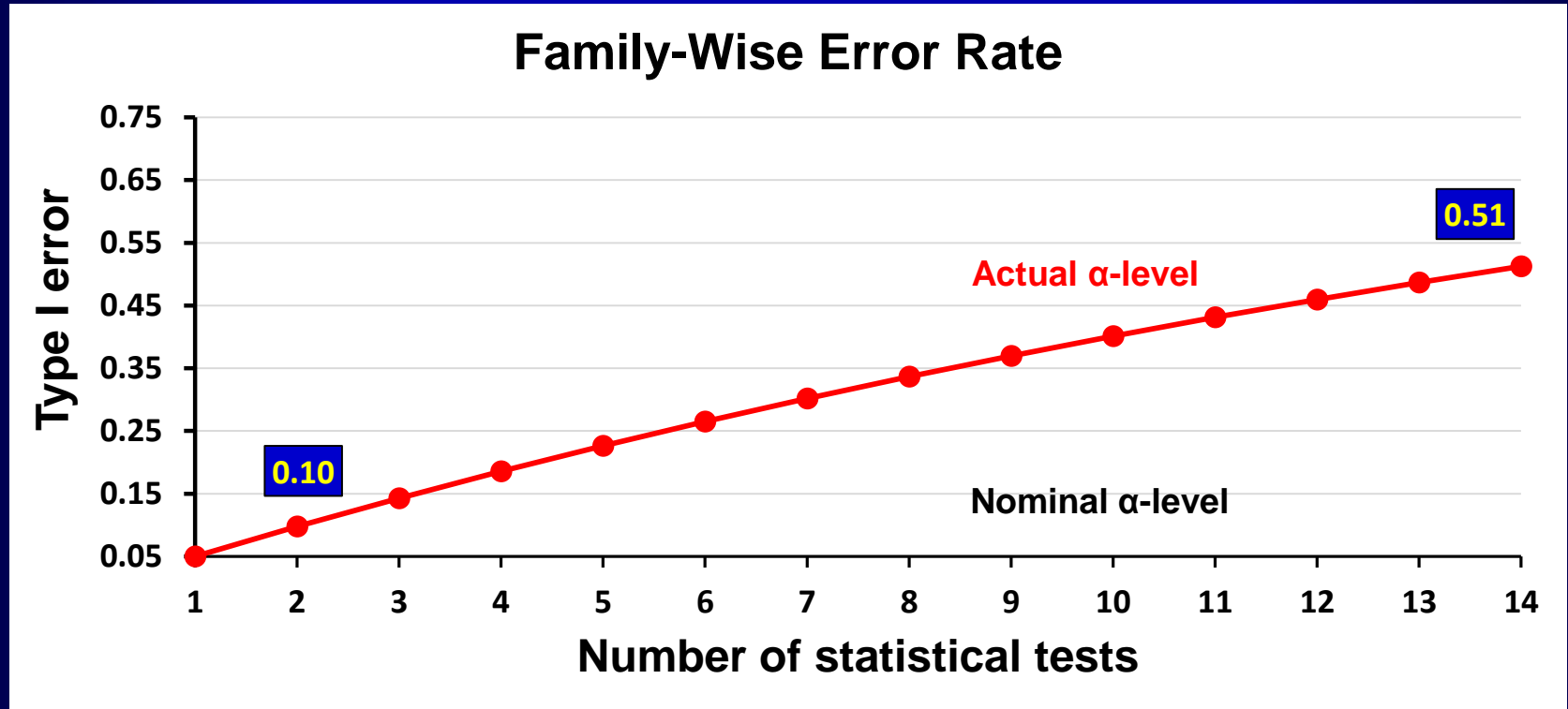


Multiple outcomes



Increasingly inflated Type I error

Multiple outcomes



Family-Wise Error Rate = probability of at least one false positive

Multiple outcomes

Alternative options:

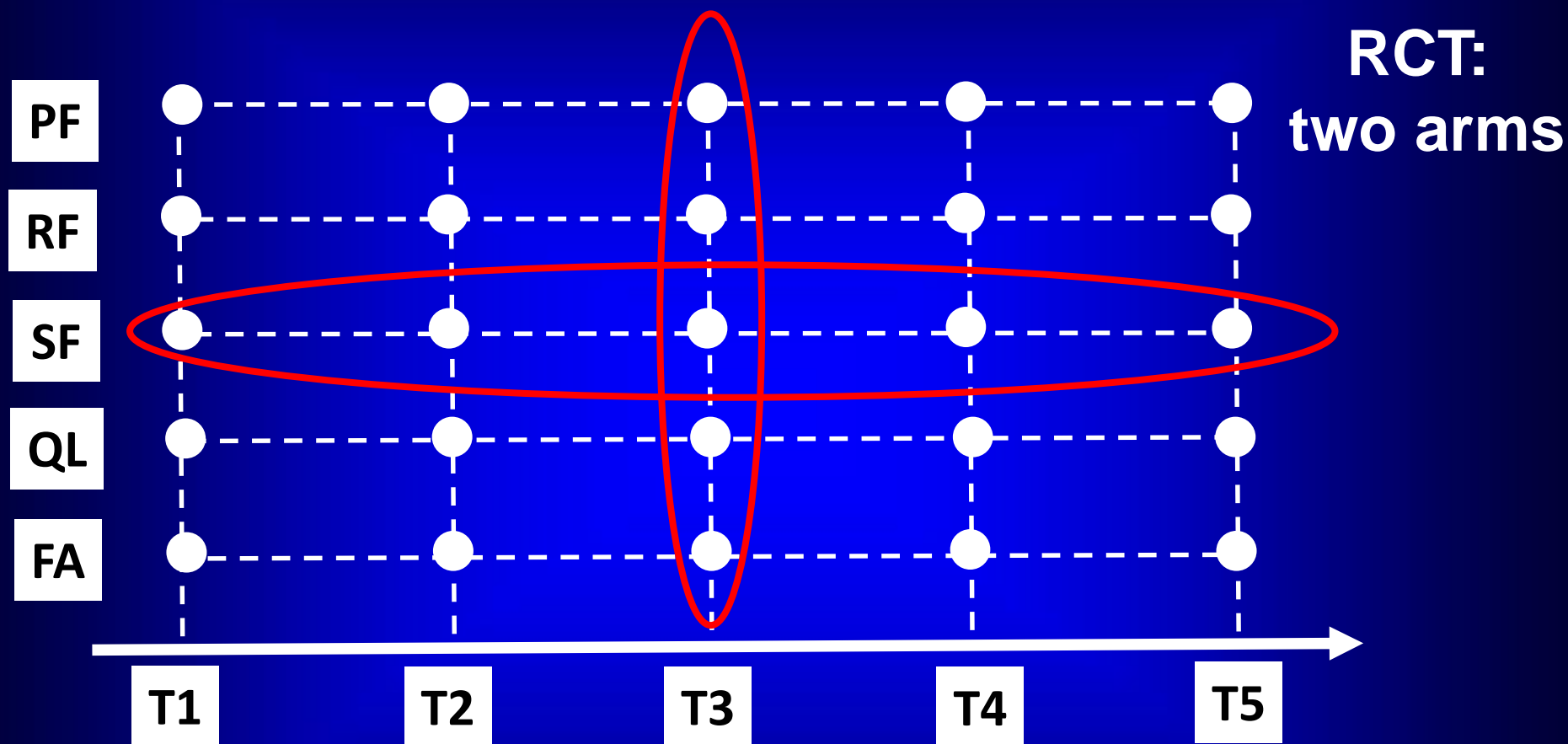
- One primary outcome
- A summary combination of outcomes
- Select k primary outcomes


Number
of
repeated
assessments




Adjust for multiple testing = limit FWER at α -level
Bonferroni, $\alpha^* = \alpha/k$

Multiple outcomes



25 pairwise t-tests  FWER = 0.72, $\alpha^* = 0.002$

5 overall F-tests  FWER = 0.23, $\alpha^* = 0.01$

Multiple outcomes

- **Plan the least possible number of statistical tests**
- **If multiple outcomes/assessments :**
adjust for multiple testing (confirmatory studies)
- **If multiple outcomes/assessments:**
do not adjust for multiple testing (exploratory/descriptive studies)
- **Whatever the choice, clearly state it in protocol/publication!**

Multiple outcomes

Multicollinearity

Table 4. Spearman Correlations Between the EORTC Scales*

EORTC Scales	PF	RF	CF	EF	SF	QL	F	P
Physical functioning (PF)								
Role functioning (RF)	0.52							
Cognitive functioning (CF)	0.40	0.41						
Emotional functioning (EF)	0.30	0.31	0.58					
Social functioning (SF)	0.52	0.66	0.49	0.45				
Global quality of life (QL)	0.55	0.61	0.56	0.52	0.63			
Fatigue (F)	-0.65	-0.75	-0.47	-0.52	-0.67	-0.71		
Pain (P)	-0.40	-0.53	-0.32	-0.42	-0.48	-0.53	0.63	
Nausea/vomiting (NV)	-0.21	-0.34	-0.19	-0.29	-0.26	-0.34	0.44	0.32

Sherman AC, Simonton S et al., Arch Otolaryngol Head Neck Surg. 2000 Apr;126(4):459-67.

Multiple outcomes

Multicollinearity in prognostic factor analysis



Incorrect model selection



**Magnitude/direction
of coefficients**

Van Steen, K., et al., Stat Med 2002

Actual multicollinearity among selected multiple scales will show in the analysis phase!

Multiple outcomes

Possible approaches

- Retain one of the collinear variables



Loss of information

- Combine collinear variables in a new summary measure



Loss of information

- Combine collinear variables in mutually orthogonal factors

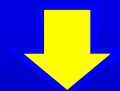


Difficult clinical interpretation

Multiple outcomes

Alternative approach

- Allows retaining multiple collinear PROs
- Useful with no *a priori* knowledge about relevant PROs
- Particularly useful with small sample sizes



Penalized methods

Multiple outcomes

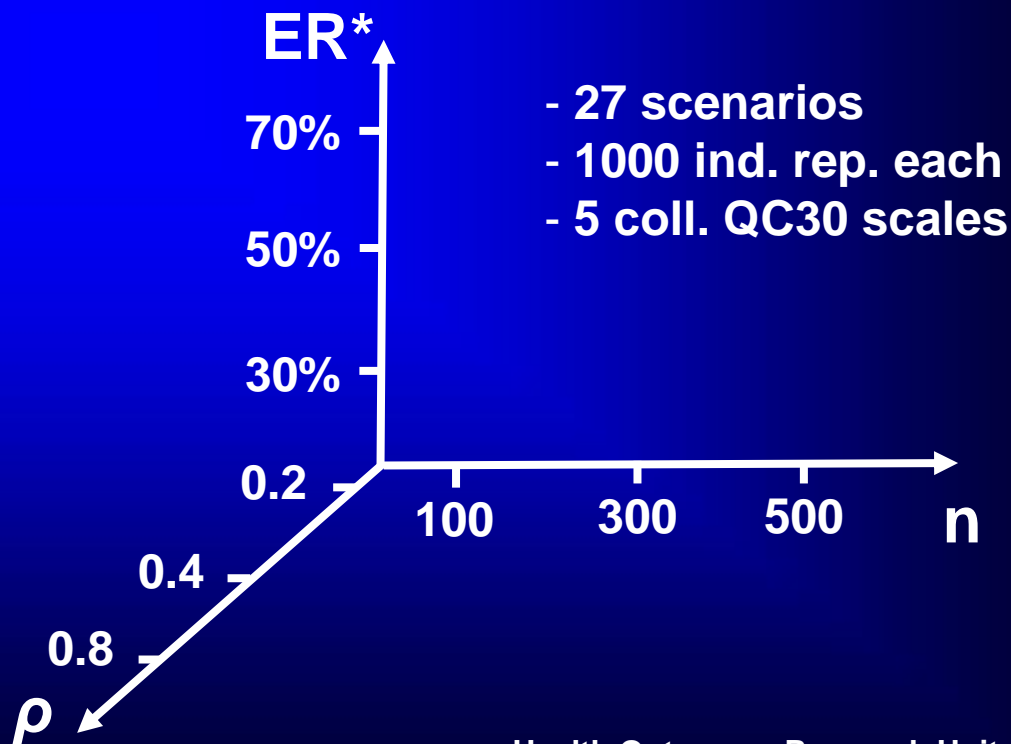
Traditional methods

1. **Cox-PV** : stepwise Cox PH - Likelihood Ratio-Test
2. **Cox-AIC** : stepwise Cox PH - Akaike Information Criterion
3. **Cox-Full** : full Cox PH

Penalized methods

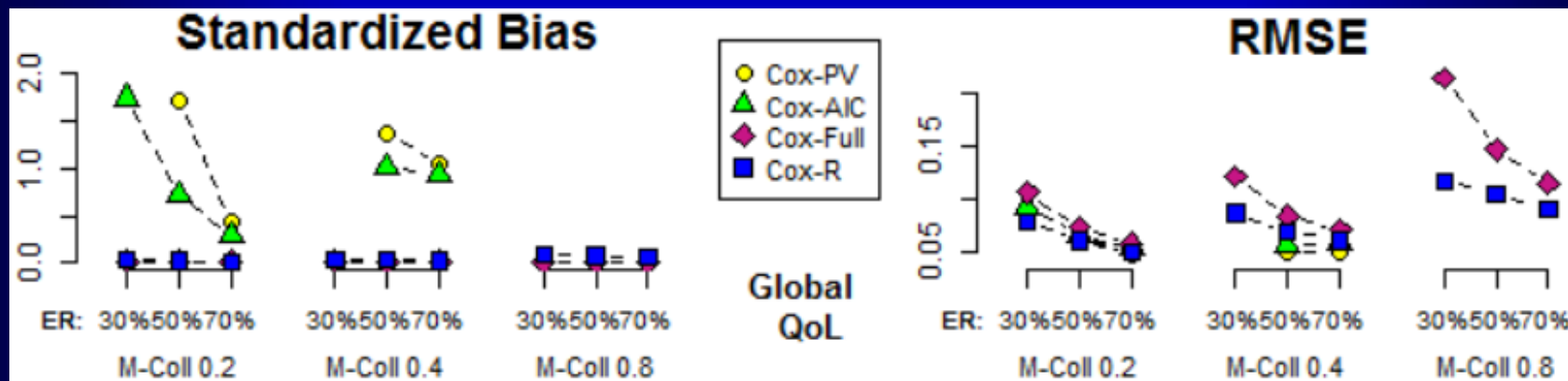
4. **Cox-R** : Ridge Cox PH
5. **Cox-EN** : Elastic-Net Cox PH

* = event rate

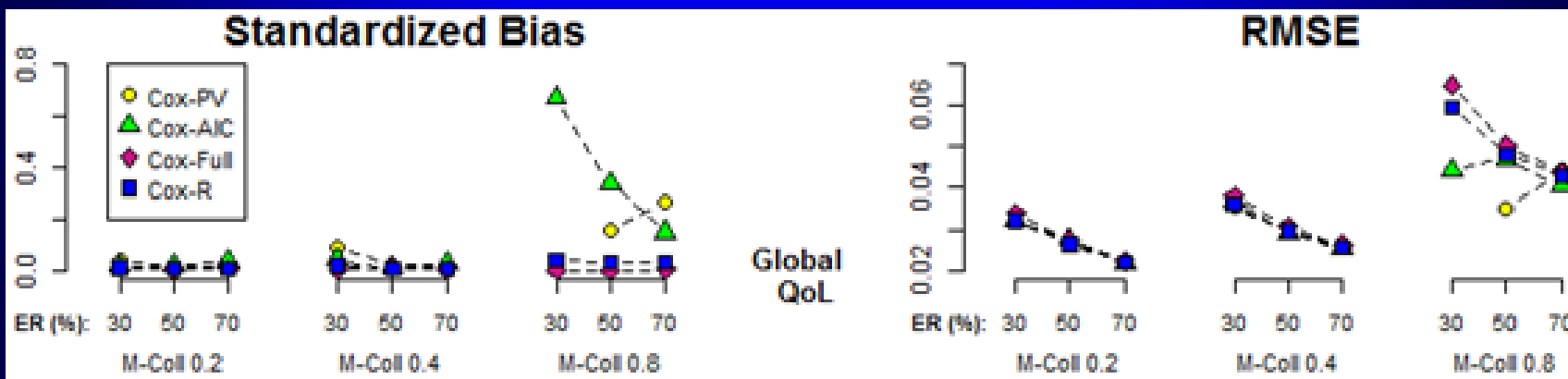


Multiple outcomes

n=100



n=500



Clinical significance

- Needed to define our primary endpoint(s)

"We performed calculation in order to detect, at least, an hazard ratio (HR) for overall survival of 1.010 for the baseline fatigue scale"

- Different types of measures



- Different thresholds for the same endpoint

Clinical significance

Different thresholds for the same endpoint

$$\Delta=10 \rightarrow ES=0.4$$

$$ES=0.5 \rightarrow \Delta=12.5$$

100 x arm

Statistic
(SD=25)

Sample size

Type I error

Power

0.05

80%

* Osoba D, et al J Clin Oncol 1998;1:139–44.

** Norman GR et al. Med Care 2003;41:582–92

Clinical significance

Cross-sectional differences

Table 4. Guidelines for Size of Cross-Sectional Differences (from meta-analysis)

Subscale	Mean Difference				Effect Size†			
	Trivial	Small	Medium	Large	Trivial	Small	Medium	Large
DI	0-3	3-7	> 7	—	0-0.1	0.1-0.4	> 0.4	—
NV	0-3	3-8	8-15	> 15	0-0.2	0.2-0.5	0.5-0.8	> 0.8
CF	0-3	3-9	9-14	> 14	0-0.2	0.2-0.4	0.4-0.7	> 0.7
DY	0-4	4-9	9-15	> 15	0-0.1	0.1-0.3	0.3-0.6	> 0.6
FI	0-3	3-10	> 10	—	0-0.1	0.1-0.4	> 0.4	—
QL	0-4	4-10	10-15	> 15	0-0.2	0.2-0.4	0.4-0.6	> 0.6
SF	0-5	5-11	11-15	> 15	0-0.2	0.2-0.4	0.4-0.6	> 0.6
SL	0-4	4-13	13-24	> 24	0-0.1	0.1-0.5	0.5-1	> 1
FA	0-5	5-13	13-19	> 19	0-0.2	0.2-0.5	0.5-0.8	> 0.8
CO	0-5	5-13	13-19	> 19	0-0.2	0.2-0.5	0.5-0.8	> 0.8
PA	0-6	6-13	13-19	> 19	0-0.2	0.2-0.5	0.5-0.8	> 0.8
PF	0-5	5-14	14-22	> 22	0-0.2	0.2-0.6	0.6-1	> 1
AP	0-5	5-14	14-23	> 23	0-0.2	0.2-0.6	0.6-1	> 1
RF	0-6	6-19	19-29	> 29	0-0.2	0.2-0.7	0.7-1.1	> 1.1

Cocks K., et al. J Clin Oncol. 2011 Jan 1;29(1):89-96.

Clinical significance

Longitudinal HRQoL deterioration

Sub-scale	Deteriorations			No difference
	Large	Medium ^a	Small	Trivial
FI	NE	<-10	-10 to -2	-2 to 3
CF	NE	<-7	-7 to -1	-1 to 3
PF	<-17	-17 to -10	-10 to -5	-5 to 2
QL	<-16	-16 to -10	-10 to -5	-5 to 5
SF	NE	<-11	-11 to -6	-6 to 3
EF	NE	<-12	-12 to -3	-3 to 6
NV	<-16	-16 to -11	-11 to -5	-5 to 3
DY	NE	<-11	-11 to -5	-5 to 2
FA	<-15	-15 to -10	-10 to -5	-5 to 4
SL	<-17	-17 to -9	-9 to -2	-2 to 5
PA	<-20	-20 to -11	-11 to -3	-3 to 5
CO	NE	<-15	-15 to -5	-5 to 4
DI	NE	<-15	-15 to -5	-5 to 3
RF	<-22	-22 to -14	-14 to -7	-7 to 6
AP	<-26	-26 to -14	-14 to -2	-2 to 7

Clinical significance

Longitudinal HRQoL improvement

Sub-scale	No difference	Improvements		
	Trivial	Small	Medium ^a	Large
FI	-2 to 3	>3	NE	NE
CF	-1 to 3	3-7	>7	NE
PF	-5 to 2	2-7	>7	NE
QL	-5 to 5	5-8	>8	NE
SF	-6 to 3	3-8	>8	NE
EF	-3 to 6	6-9	>9	NE
NV	-5 to 3	3-9	>9	NE
DY	-5 to 2	2-9	>9	NE
FA	-5 to 4	4-9	>9	NE
SL	-2 to 5	5-9	>9	NE
PA	-3 to 5	5-9	9-14	>14
CO	-5 to 4	4-10	>10	NE
DI	-5 to 3	3-11	>11	NE
RF	-7 to 6	6-12	>12	NE
AP	-2 to 7	7-13	>13	NE

Clinical significance

Minimally Clinical Important Difference (MCID)

	Physical Functioning	Role Functioning	GHS/ QoL	Diarrhea	Insomnia
CS/LGT*	10	10	10	10	10
CS †	5-14	6-19	4-10	3-7	4-13
LGT-Impr. #	2-7	6-12	5-8	3-11	5-9
LGT-Deter. #	5-10	7-14	5-10	5-15	2-9

Abbreviations: GHS, Global, Health Status, CS, Cross sectional, LGT, Longitudinal, Impr.=improvement, Det.=deterioration.

* Osoba D, et al J Clin Oncol 1998;1:139–44.

† Cocks K., et al. J Clin Oncol. 2011 Jan 1;29(1):89-96.

Cocks K., et al. Eur J Cancer. 2012 Jul;48(11):1713-21.

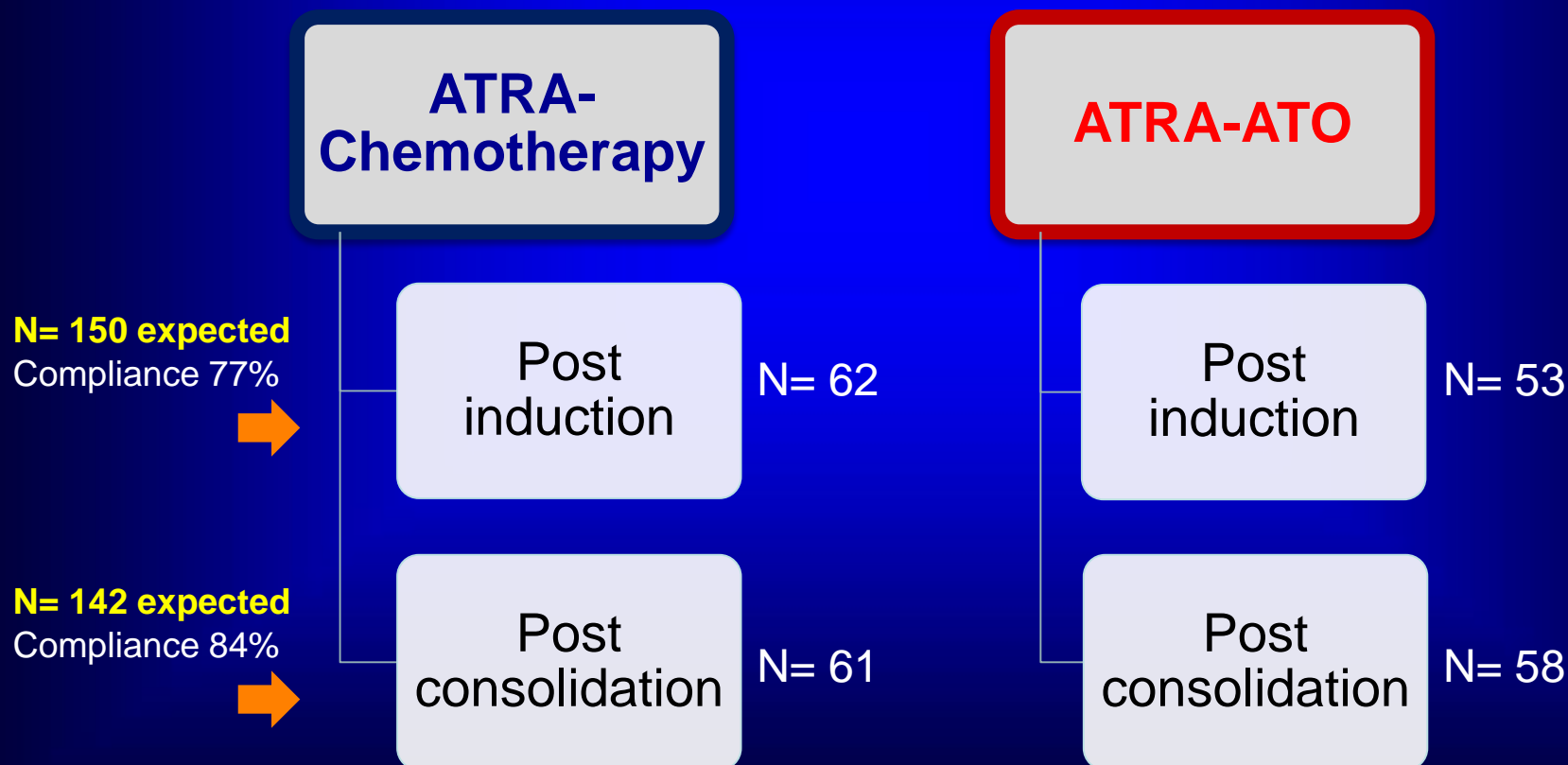
Results and QoL COMPLIANCE

Between October 2007 and September 2010, **162 patients** were enrolled.

Genetic tests excluded a diagnosis of PML/RARA-positive APL in 3 patients. Three of 159 patients with genetically proven APL did not start allocated treatment.

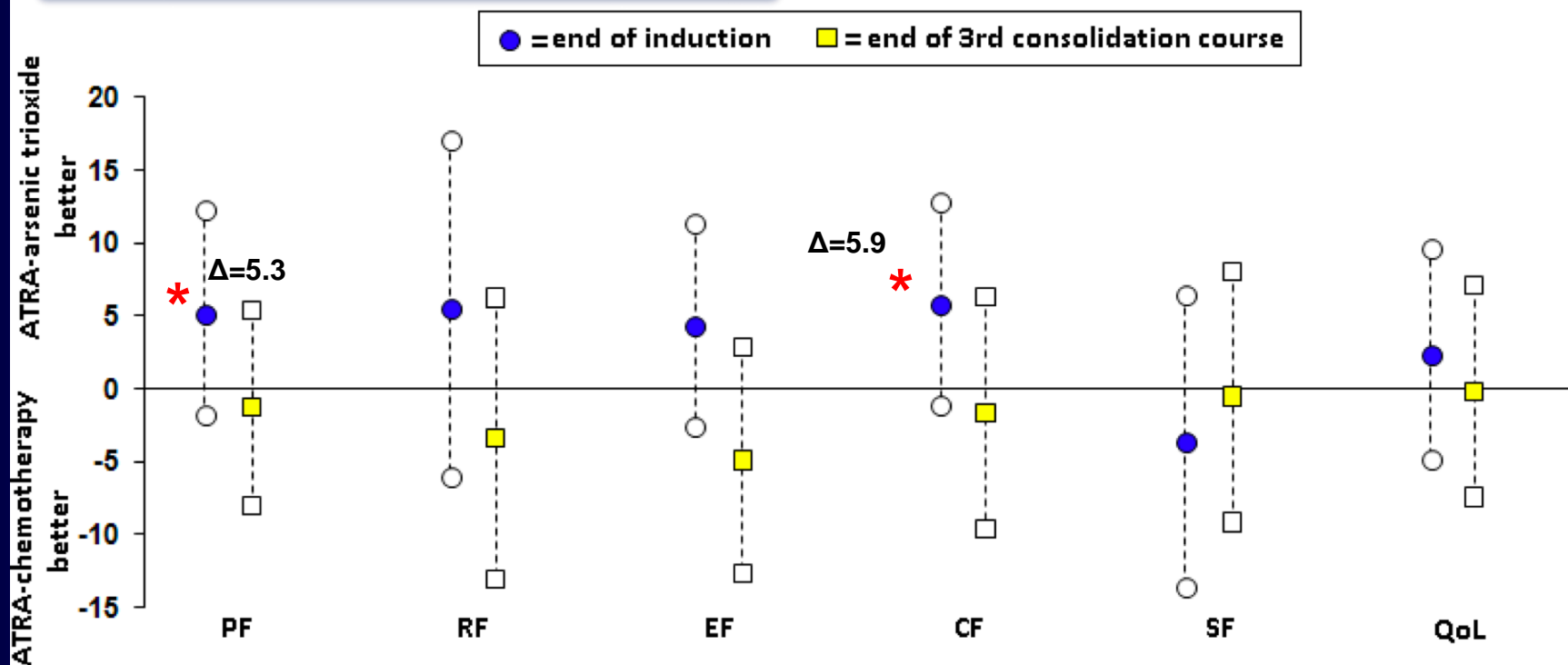
156 patients

who received at least one dose of the assigned therapy after randomization.



Estimated differences in EORTC QLQ-C30 mean scores and 95% CIs between ATRA-arsenic trioxide and ATRA-chemotherapy arms at the end of induction therapy and third consolidation course.

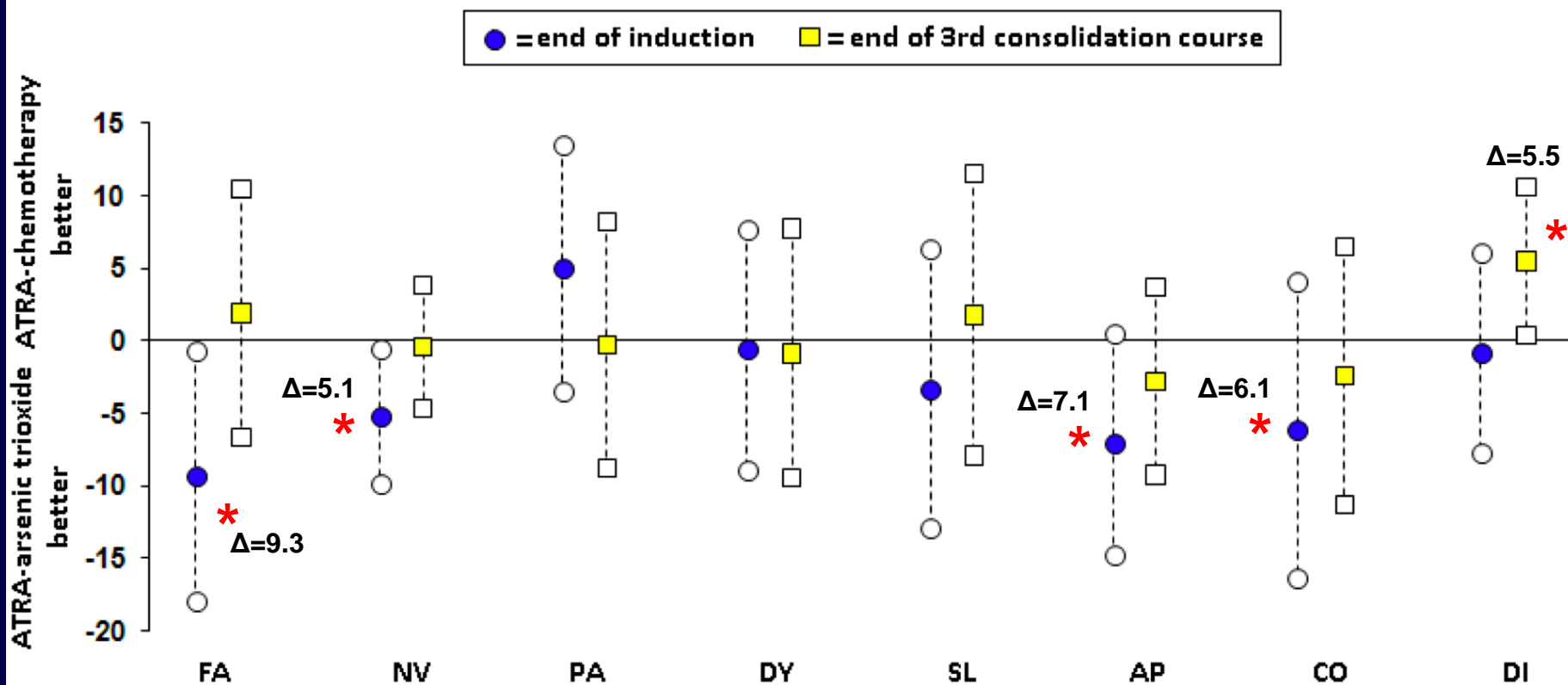
Functional aspects / Global QoL



* Clinically meaningful difference; (based on Cocks K, et al., J Clin Oncol. 2011;29:89-96).

Estimated differences in EORTC QLQ-C30 mean scores and 95% CIs between ATRA-arsenic trioxide and ATRA-chemotherapy arms at the end of induction therapy and third consolidation course.

Symptoms



* Clinically meaningful difference; (based on Cocks K, et al., J Clin Oncol. 2011;29:89-96).

Clinical significance

1. The method used to estimate clinical significance of PROs must be scientifically sound.
2. The 0.5 SD is a conservative estimate of clinical significance
3. Using 0.5 SD is fine if no more specific thresholds are available for the questionnaire/scale
4. Future development is desirable of specific thresholds for clinical significance, according to questionnaire, scale and setting.

THANKS

f.cottone@gimema.it