Leids Universitair
Medisch Centrum

On population diversity in clinical trials
A few remarks

Prof. Olaf M. Dekkers

Dept clinical epidemiology & internal medicine LUMC

Thinking about population diversity in trials

Design

- What populations to include?

Analysis

- Subgroup analysis

Interpretation

- Are results generalizable to other populations?
- (applicability)

Can we assume (or not) that treatment effects are stable over populations?

Diversity at design stage

Most trials display diversity with regard to

- Age
- Sex
- Ethnicity
- Country

Even despite eligibility criteria being strict

Diversity at design stage

Most trials display diversity with regard to

- Age
- Sex
- Ethnicity
- Country

Even despite eligibility criteria being strict

In most trials it is assumed that treatment effects are stable over ***included*** populations

Think of this from the perspective of the designer:
If subgroups are expected not to show an effect they are likely not included (power, costs)

Diversity at interpretation stage

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

ORIGINAL ARTICLE

Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism

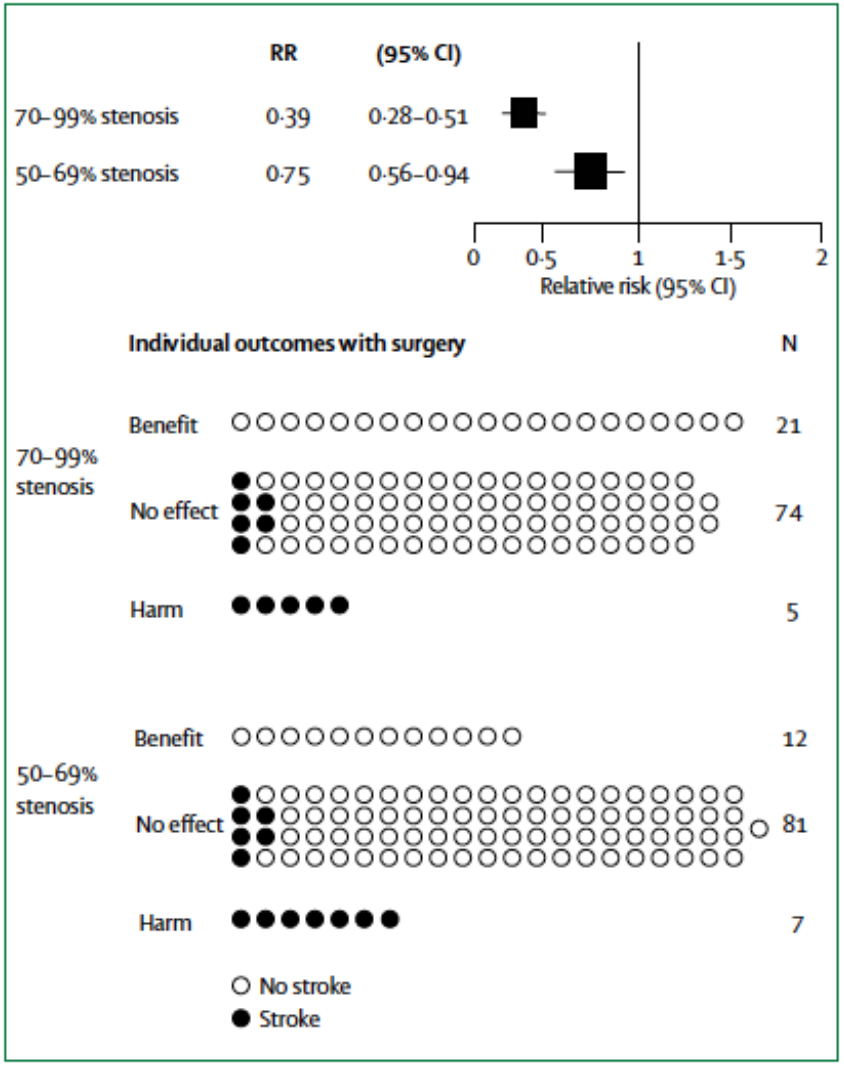
David J. Stott, M.B., Ch.B., M.D., Nicolas Rodondi, M.D., Patricia M. Kearney, M.D., Ph.D., Ian Ford, Ph.D., Rudi G.J. Westendorp, M.D., Ph.D., Simon P. Mooijaart, M.D., Ph.D., Naveed Sattar, F.Med.Sci., Carole E. Aubert, M.D., Drahomir Aujesky, M.D., Douglas C. Bauer, M.D., Christine Baumgartner, M.D., Manuel R. Blum, M.D., John P. Browne, Ph.D., Stephen Byrne, Ph.D., Tinh-Hai Collet, M.D., Olaf M. Dekkers, M.D., Ph.D., Wendy P.J. den Elzen, Ph.D., Robert S. DuPuy, M.D., Graham Ellis, M.D., Martin Feller, M.D., Carmen Floriani, M.D., Kirsty Hendry, Ph.D., Caroline Hurley, M.P.H., J. Wouter Jukema, M.D., Ph.D., Sharon Kean, Maria Kelly, M.Pharm., Danielle Krebs, Ph.D., Peter Langhorne, M.D., Ph.D., Gemma McCarthy, M.P.H., Vera McCarthy, Ph.D., Alex McConnachie, Ph.D., Mairi McDade, B.Sc., R.G.N., Martina Messow, Ph.D., Annemarie O'Flynn, Ph.D., David O'Riordan, M.Pharm., Rosalinde K.E. Poortvliet, M.D., Ph.D., Terence J Quinn, M.D., Ph.D., Audrey Russell, M.M.Sc., Carol Sinnott, Ph.D., Jan W.A. Smit, M.D., Ph.D., H. Anette Van Dorland, Ph.D., Kieran A. Walsh, M.Pharm., Elaine K. Walsh, M.B., B.Ch., B.A.O., Torquil Watt, M.D., Robbie Wilson, M.Sc., and Jacobijn Gussekloo, M.D., Ph.D., for the TRUST Study Group*

April 3, 2017 | DOI: 10.1056/NEJMoa1603825

Share:

Abstract | Article | References | Metrics

The problem of average effects



Huge discrepancy between average population effects and effects on individual levels

A trial: basic ‘emotions’ for subgroup analysis

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

ORIGINAL ARTICLE

Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism

David J. Stott, M.B., Ch.B., M.D., Nicolas Rodondi, M.D., Patricia M. Kearney, M.D., Ph.D., Ian Ford, Ph.D., Rudi G.J. Westendorp, M.D., Ph.D., Simon P. Mooijaart, M.D., Ph.D., Naveed Sattar, F.Med.Sci., Carole E. Aubert, M.D., Drahomir Aujesky, M.D., Douglas C. Bauer, M.D., Christine Baumgartner, M.D., Manuel R. Blum, M.D., John P. Browne, Ph.D., Stephen Byrne, Ph.D., Tinh-Hai Collet, M.D., Olaf M. Dekkers, M.D., Ph.D., Wendy P.J. den Elzen, Ph.D., Robert S. DuPuy, M.D., Graham Ellis, M.D., Martin Feller, M.D., Carmen Floriani, M.D., Kirsty Hendry, Ph.D., Caroline Hurley, M.P.H., J. Wouter Jukema, M.D., Ph.D., Sharon Kean, Maria Kelly, M.Pharm., Danielle Krebs, Ph.D., Peter Langhorne, M.D., Ph.D., Gemma McCarthy, M.P.H., Vera McCarthy, Ph.D., Alex McConnachie, Ph.D., Mairi McDade, B.Sc., R.G.N., Martina Messow, Ph.D., Annemarie O'Flynn, Ph.D., David O'Riordan, M.Pharm., Rosalinde K.E. Poortvliet, M.D., Ph.D., Terence J Quinn, M.D., Ph.D., Audrey Russell, M.M.Sc., Carol Sinnott, Ph.D., Jan W.A. Smit, M.D., Ph.D., H. Anette Van Dorland, Ph.D., Kieran A. Walsh, M.Pharm., Elaine K. Walsh, M.B., B.Ch., B.A.O., Torquil Watt, M.D., Robbie Wilson, M.Sc., and Jacobijn Gussekloo, M.D., Ph.D., for the TRUST Study Group*

April 3, 2017 | DOI: 10.1056/NEJMoa1603825

Share:

Abstract | Article | References | Metrics

What is a good rationale for subgroup analysis?

To avoid false positive results from subgroup analysis, a good rationale is said to be a prerequisite

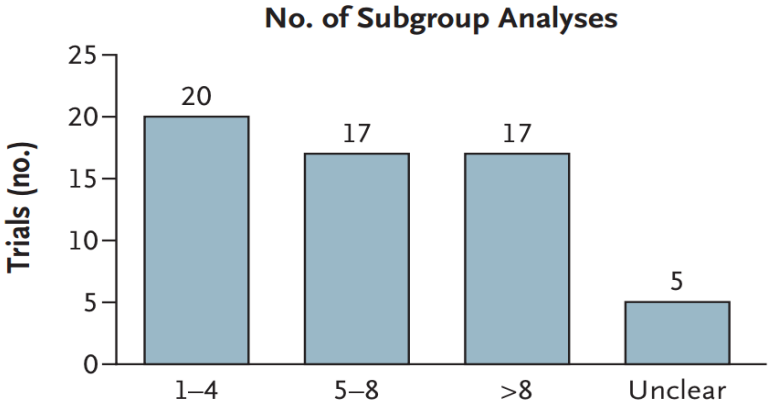
Rationale may be biological or clinical (and not political)

Mind that the rationale should be about effect modification, not about why a treatment works in subgroup X or Y

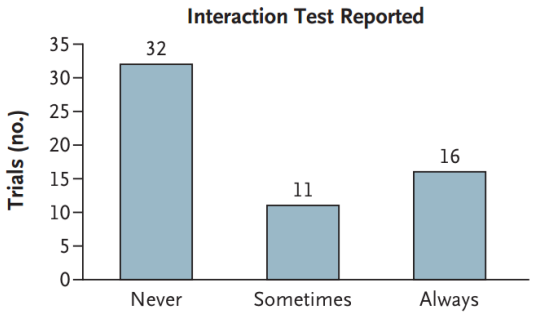
Example:
Olanzapine vs risperidone in psychosis for endpoint prolactin levels (clinical rationale)

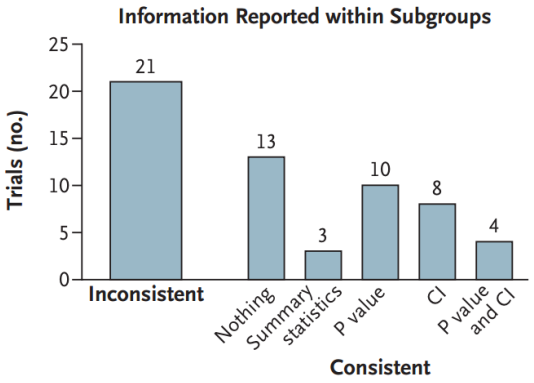
Diversity: explore or ignore?

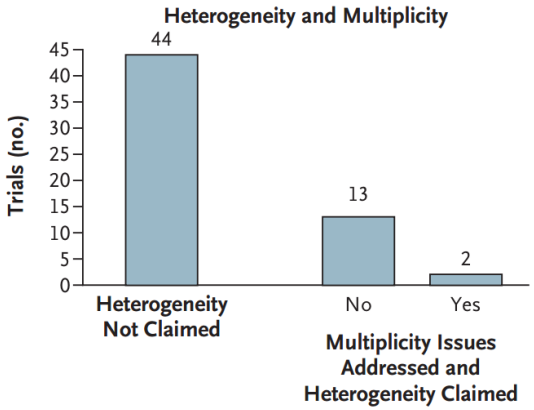
Subgroup analysis reported in 59/97 in trials

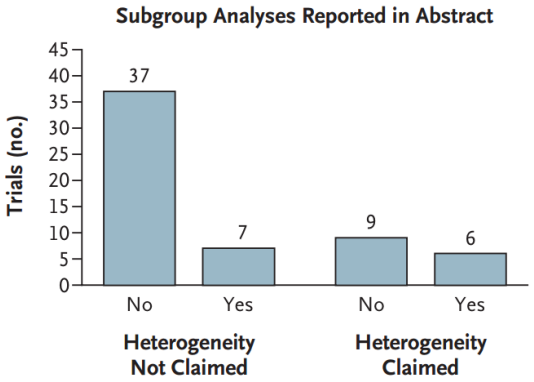


Subgroup analysis reported in 59/97 in trials



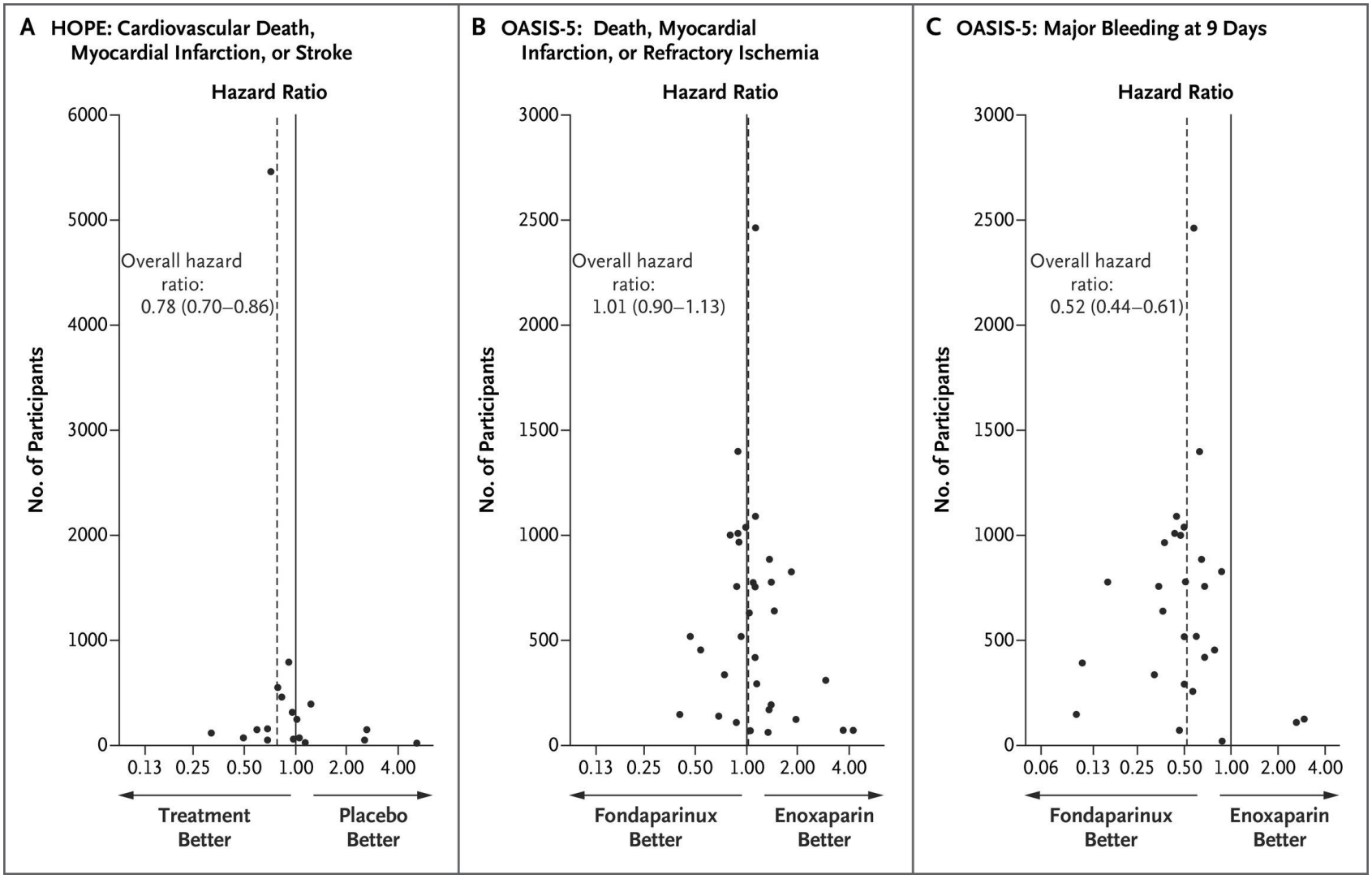






Analysis often inadequately performed/reported

Population diversity (country)



Probability of false positive results

If p=0.05 every true H0 has a probability of being significant of 5%

Number of tests	Probability of no false positive test	
1	0.95*1	95%
2	0.95*2	90%
3	0.95*3	86%
4	0.95*4	81%
5	0.95*5	77%
10	0.95*10	60%
20	0.95*20	36%

A closer look at diversity: sex

Table 1 | Summary results for proportion of statistically significant sex-treatment interactions based on different eligible criteria

Eligibility criteria	No of topics (No of trials)	No (%) of statistically significant sex-treatment interactions
All	109 (311)	8/109 (7)
Only topics with data for both men and women	96 (162)	6/96 (6)
Only topics with >1 RCT and at least one RCT with data for both men and women*	39 (209)	3/39 (8)
As above, but excluding RCTs with data on only one sex*	39 (106)	4/39 (10)
One topic per review (most inclusive topic with the most data)	41† (164)	4/41 (10)
One topic per treatment comparison (most inclusive topic with the most data)	61‡ (194)	7/61 (11)

Corroboration of subgroup findings (sample of 46 subgroup claims)

Table 3. Five Subgroup Findings With Full Corroboration Attempts

Characteristics of the Subgroup Findings (Index Articles)					Results of the Corroboration Attempts[a]
Comparison (Year)	Subgroups	Population Characteristics	Outcome (Primary)	Index Article *P* Value for Interaction	*P* Value for Interaction Corroboration Attempt
Supportive expressive group therapy vs control (2007)[22]	Estrogen receptor status negative vs positive	Women with confirmed metastatic or locally recurrent breast cancer	Survival (yes)	.002	.71
Standard care vs standard care without intravenous cooling (2007)[23]	Patients with initial ventricular fibrillation vs no ventricular fibrillation	Patients aged ≥18 y, resuscitated by paramedics from nontraumatic, out-of-hospital cardiac arrest	Awakening (no)	.046[b]	No *P* value provided, no evidence of subgroup difference
			Discharge alive from hospital (no)	.048[b]	.52[b]
Dexamethasone sodium phosphate vs placebo (2007)[24]	Patients with confirmed bacterial meningitis vs probable meningitis	Patients aged ≥14 y with suspected bacterial meningitis	Risk of death at 1 mo (yes)	.01[c]	.23
N-terminal brain natriuretic peptide-guided treatment vs symptom-guided treatment (2009)[25]	Patients aged 60-74 vs ≥75 y	Patients aged ≥60 y with systolic heart failure, New York Heart Association class II or greater, prior hospitalization for heart failure within 1 y, and N-terminal brain natriuretic peptide level ≥2 times the upper limit of normal	Mortality (no)	.01	.22

Take the plus, forget the minus?

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism

D.J. Stott, N. Rodondi, P.M. Kearney, I. Ford, R.G.J. Westendorp, S.P. Mooijaart, N. Sattar, C.E. Aubert, D. Aujesky, D.C. Bauer, C. Baumgartner, M.R. Blum, J.P. Browne, S. Byrne, T.-H. Collet, O.M. Dekkers, W.P.J. den Elzen, R.S. Du Puy, G. Ellis, M. Feller, C. Floriani, K. Hendry, C. Hurley, J.W. Jukema, S. Kean, M. Kelly, D. Krebs, P. Langhorne, G. McCarthy, V. McCarthy, A. McConnachie, M. McDade, M. Messow, A. O'Flynn, D. O'Riordan, R.K.E. Poortvliet, T.J Quinn, A. Russell, C. Sinnott, J.W.A. Smit, H.A. Van Dorland, K.A. Walsh, E.K. Walsh, T. Watt, R. Wilson, and J. Gussekloo, for the TRUST Study Group*

Subgroup analysis according to
- TSH level
- Sex
- Symptom score
- Age

Table 2. Outcomes at 12 Months and Extended Follow-up.*

Variable	Baseline		At 12 Mo			
	Placebo (N=369)	Levothyroxine (N=368)	Placebo (N=320)	Levothyroxine (N=318)	Difference (95% CI)	P Value
Thyrotropin — mIU/liter	6.38±2.01	6.41±2.01	5.48±2.48	3.63±2.11	−1.92 (−2.24 to −1.59)	<0.001
Median (IQR)	5.76 (5.10 to 6.94)	5.70 (5.12 to 6.83)	4.90 (3.91 to 6.46)	3.16 (2.45 to 4.22)	—	—
Primary outcomes:						
Hypothyroid Symptoms score	16.9±17.9	17.5±18.8	16.7±17.5	16.6±16.9	0.0 (−2.0 to 2.1)	0.99
Tiredness score	25.5±20.3	25.9±20.6	28.6±19.5	28.7±20.2	0.4 (−2.1 to 2.9)	0.77

The interpretation of subgroup analyses I

Subgroup	Events/patients Surgical	Medical	RR	95% CI
Sex				
Male	92/890	172/784	0·46	0·41–0·51
Female	59/436	55/346	0·84	0·63–1·12

There seems effect modification by sex

Conclusion:

It works in both men and women?

It works in men and not in women?

It works in men and to a lesser extend in women?

The interpretation of subgroup analyses II

In a RCT treatment is randomized, balance of prognostic variable expected (not guarantee)

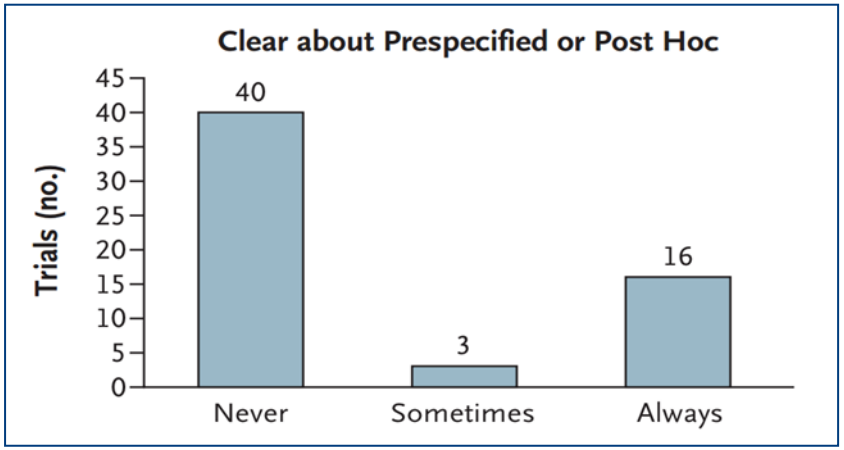
Subgroup	Events/patients		RR	95% CI
	Surgical	Medical		
Sex				
Male	92/890	172/784	0·46	0·41–0·51
Female	59/436	55/346	0·84	0·63–1·12

There seems effect modification by sex

But: characteristics of subgroups are not randomized

Confounding may exist (blood pressure, smoking status), thus subgroup analyses have no simple causal interpretation

Does prespecification matter?



Why prespecification?

Method to prevent data dredging and selective reporting

But:

Truth itself does not care whether the hypothesis was prespecified

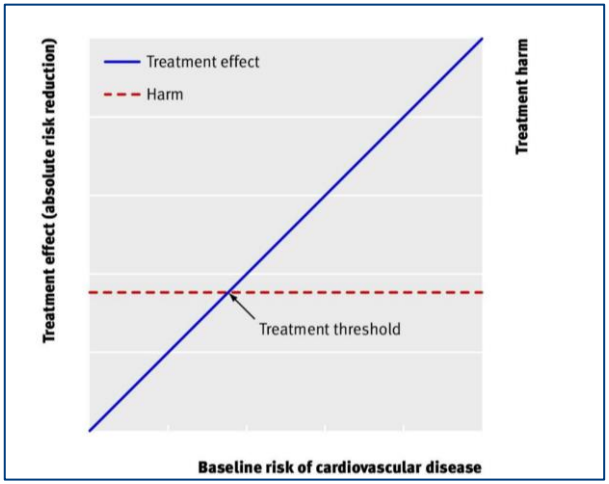
False positive results are not prevented by prespecification,
and design is not adapted (it maybe should)

So, selective reporting is the issue

Effect modification: the problem of scale

Trial with stable RR over subpopulations

Absolute risk 20%	• RR 0.8 • NNT 25
Absolute risk 10%	• RR 0.8 • NNT 50
Absolute risk 5%	• RR 0.8 • NNT 100



Preliminary conclusions

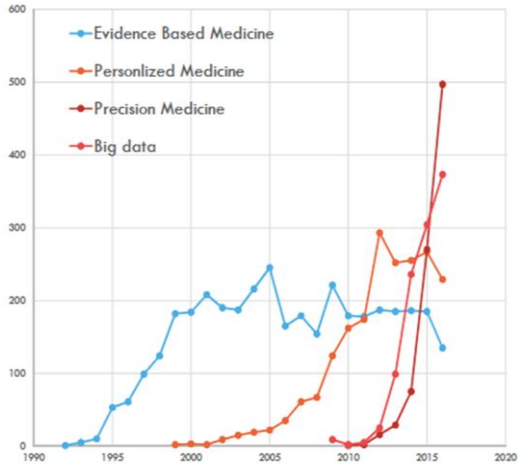
- For most trials not much effect modification is expected at the design stage
- Subgroup analyses run the risk of false-positive results
- Prespecification is important to prevent selective reporting
- Empirical evidence does not suggest much true effect modification
- The interpretation is not straightforward

Personalized medicine: diversity in ultimo?

Personalized Medicine

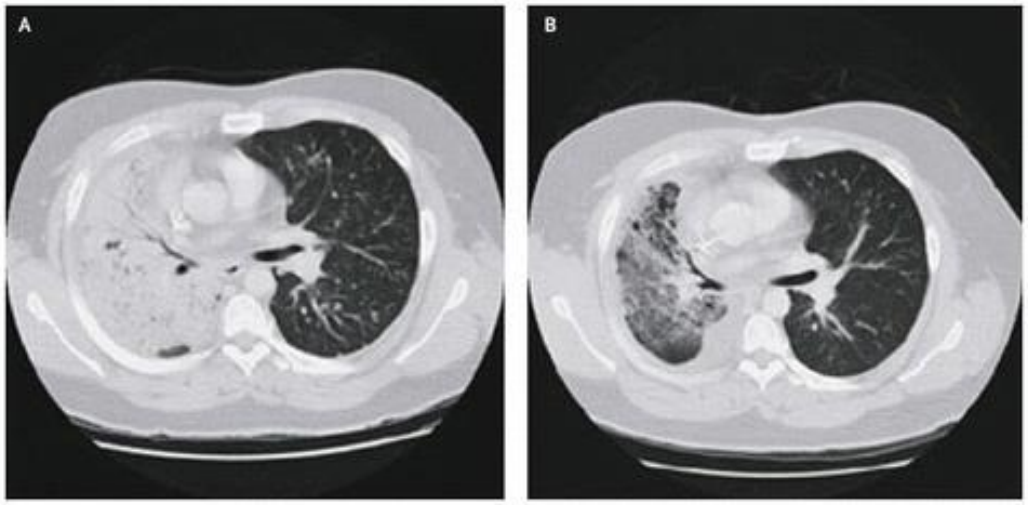
"Doctors have always recognized that every patient is unique, and doctors have always tried to tailor their treatments as best they can to individuals. You can match a blood transfusion to a blood type — that was an important discovery. What if matching a cancer cure to our genetic code was just as easy, just as standard? What if figuring out the right dose of medicine was as simple as taking our temperature?"
- President Obama, January 20, 2015

TITLES IN PUBMED

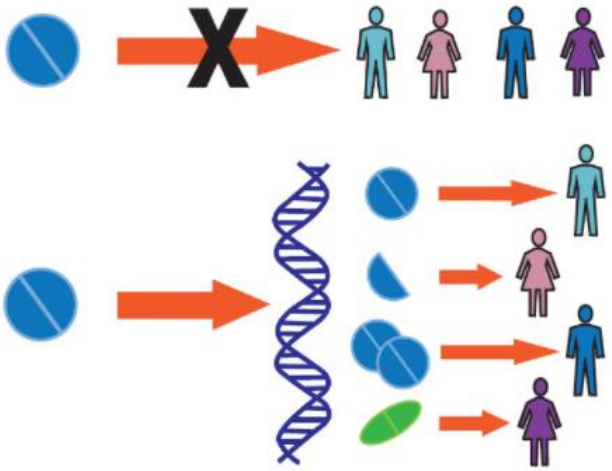


The personalized medicine principle

Selected Genomic Biomarkers.		
Biomarker	**Disease**	**Drug**
c-kit	Gastrointestinal stromal tumor	Imatinib mesylate
CCR5	Human immunodeficiency virus	Maraviroc
Cytochrome P-450 variants	Various disorders	Warfarin, voriconazole
EGFR	Non–small-cell lung cancer	Erlotinib



How to assess efficacy PM?



Two options

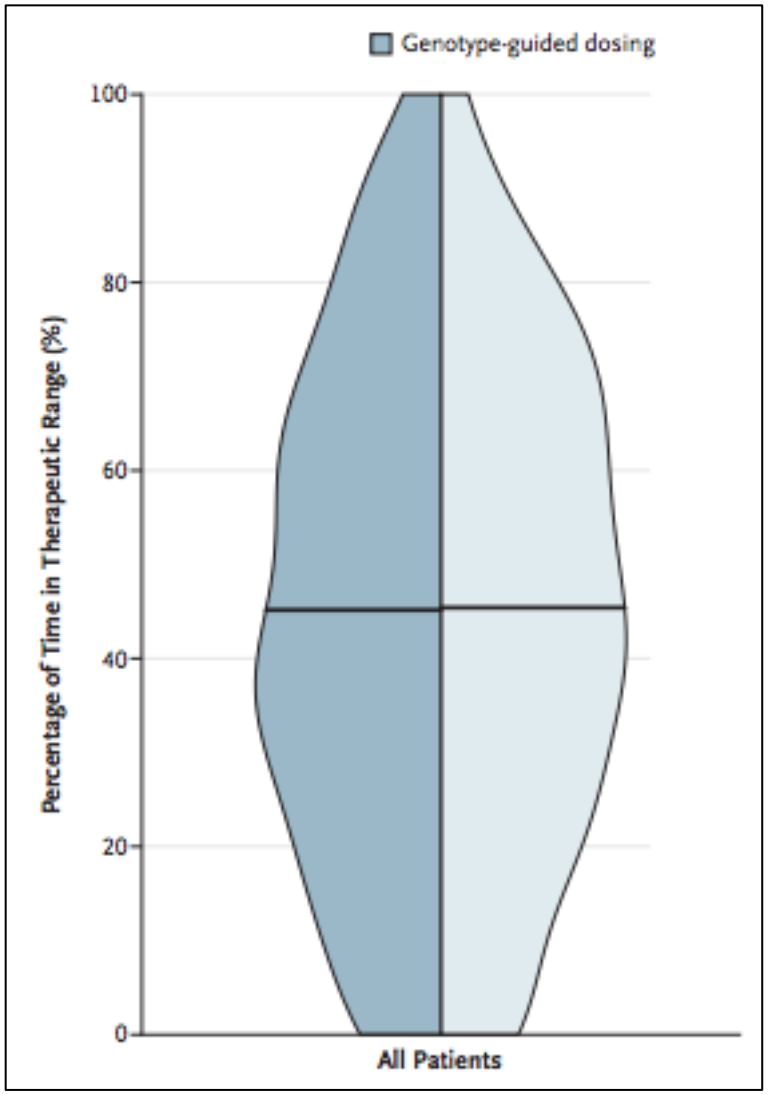
1. Subgroup analysis in trial according to gene/receptor polymorphism
2. Randomize information
3. Perfect prediction

Randomize information

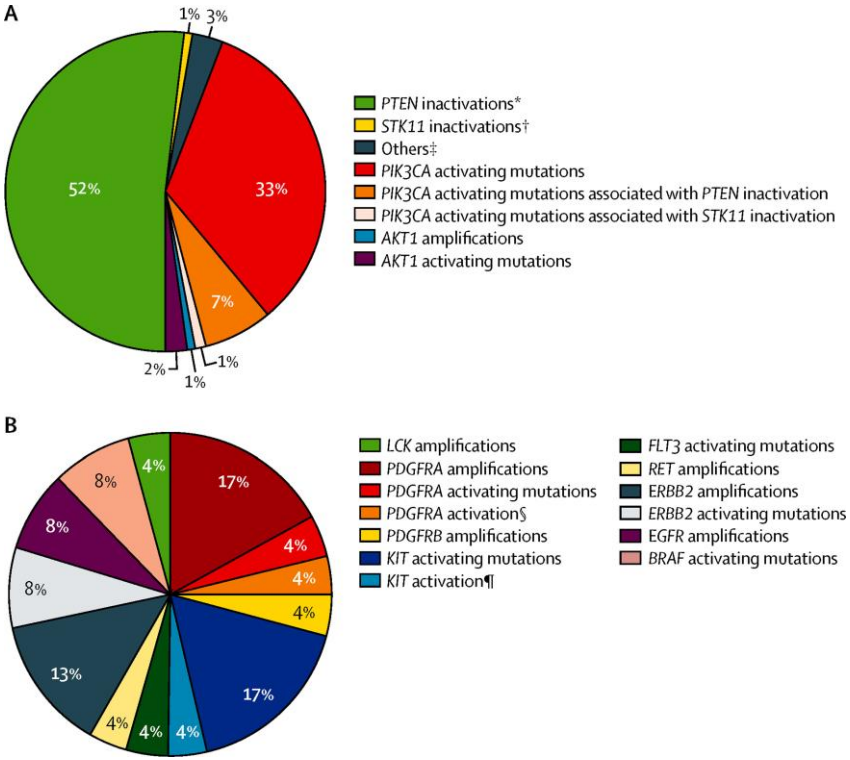
The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 DECEMBER 12, 2013 VOL. 369 NO. 24

A Pharmacogenetic versus a Clinical Algorithm for Warfarin Dosing



Randomize information



The molecular profile of a tumour was established with large-scale genomic testing. We only included patients for whom a molecular alteration was identified within one of three molecular pathways (hormone receptor, PI3K/AKT/mTOR, RAF/MEK), which could be matched to one of ten regimens including 11 available molecularly targeted agents (erlotinib, lapatinib plus trastuzumab, sorafenib, imatinib, dasatinib, vemurafenib, everolimus, abiraterone, letrozole, tamoxifen).

We randomly assigned these patients (1:1) to receive a matched molecularly targeted agent (experimental group) or treatment at physician's choice (control group)

Randomize information

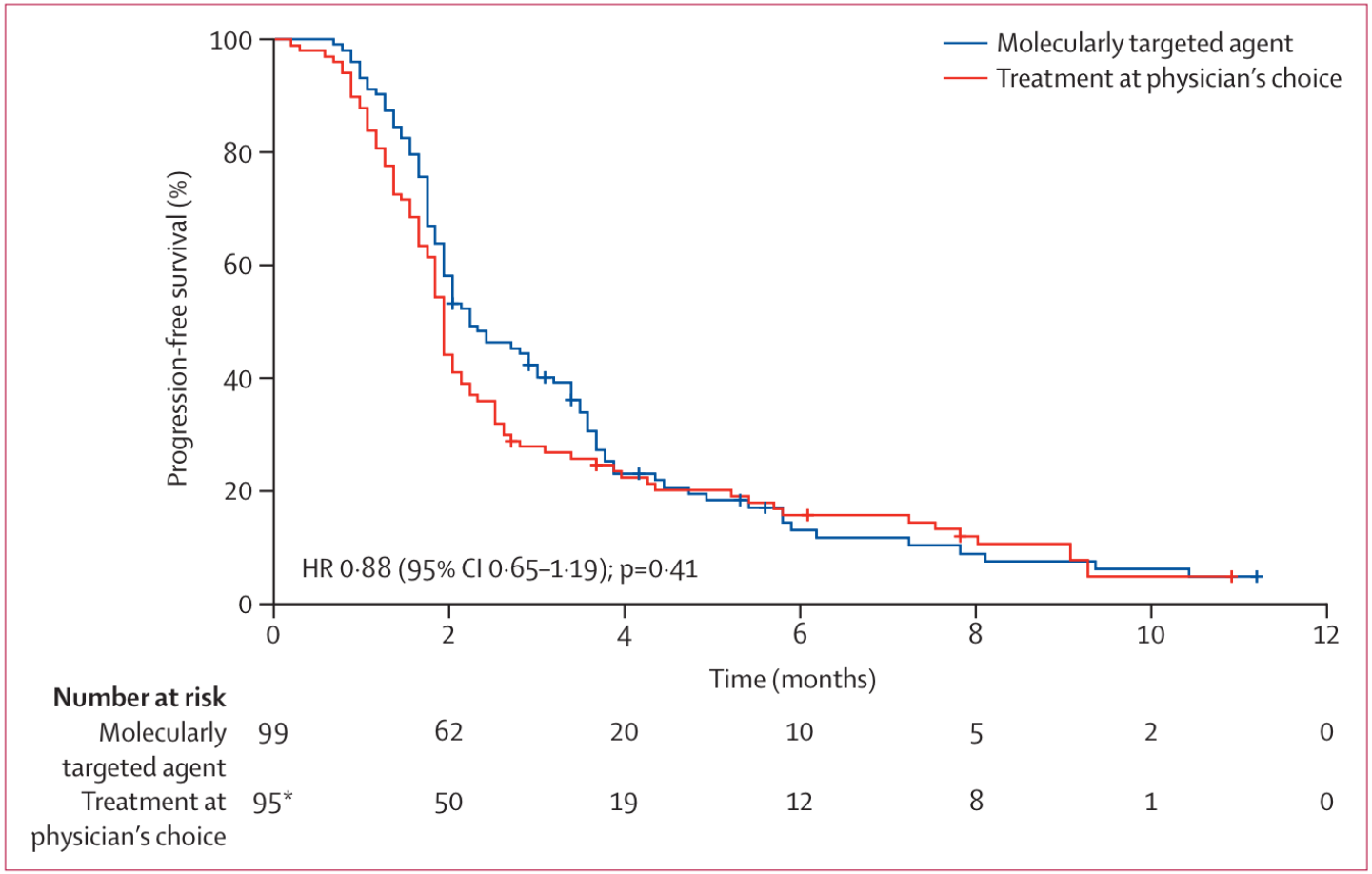
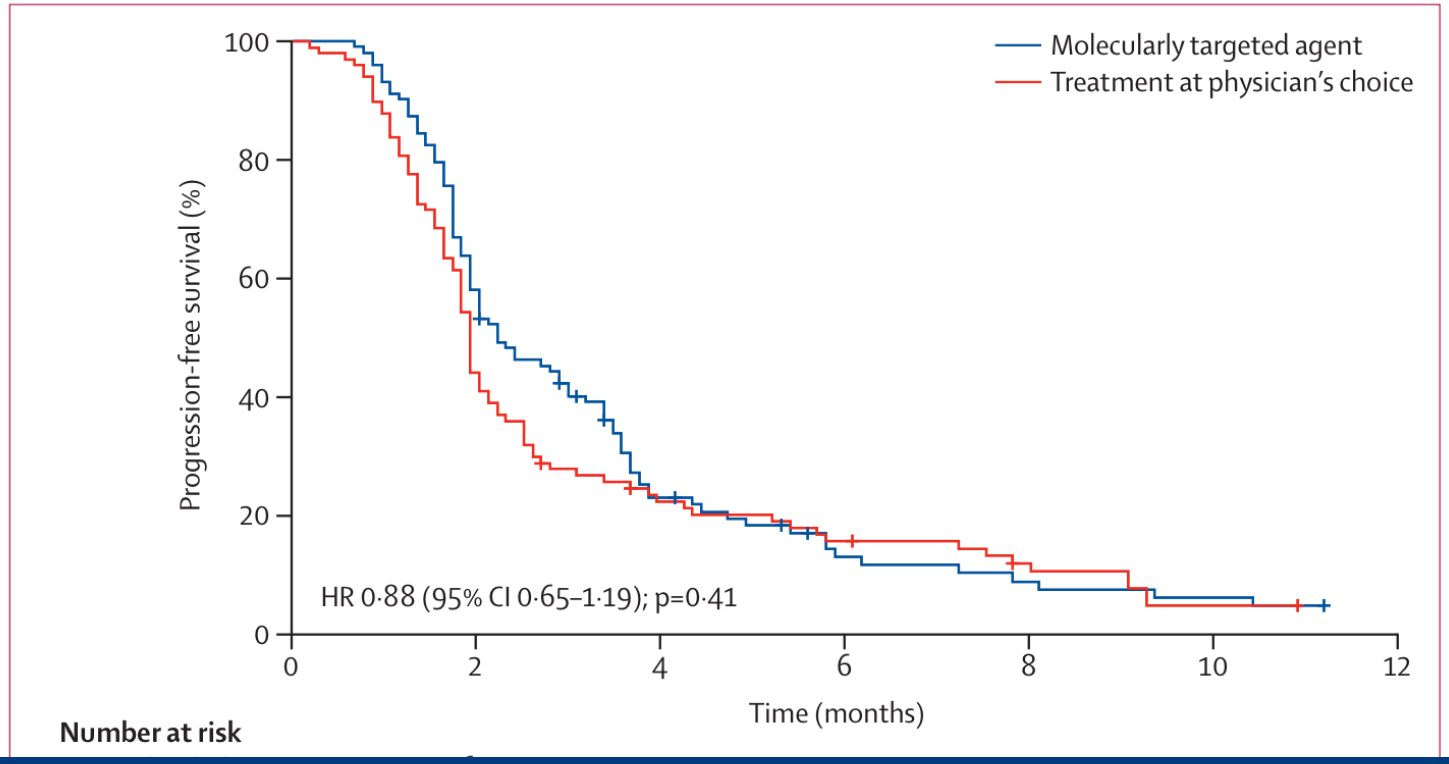


Figure 3: **Progression-free survival**

*One patient had a follow-up of zero days so is not shown here.

Randomize information



Correspondence:
In the SHIVA trial,1 28% of patient had two or more molecular alterations. Was the presence of additional harmful genomic alterations (polymutant cancers) associated with a shorter duration of response than cancers with only one identified molecular alteration (oligomutant cancers)?

To conclude

- A priori not much effect modification in trials is expected
- The main problems regarding subgroup analyses are the risk of false positive results and the difficult interpretation
- No effect modification on a RR scale translates into different NNT when generalizing evidence
- Personalized medicine is not the ultimate answer to the problem of subgroup analysis as it still requires an …. RCT