

Subgroup Analyses: Important, Infuriating and Intractable

The views expressed herein are not necessarily those of MHRA, EMA, EMA committees or their working parties.

Rob Hemmings
Statistics Unit Manager, MHRA
CHMP member
Chair, CHMP Scientific Advice Working Party
Biostatistics Working Party Member

Aims



- To give additional context to statistical discussions
- Illustrate regulatory thinking through discussion
- Provoke thought

Why is this a good topic for discussion?

- Quote from consultation *“Analyses of subgroups cause some confusion and muddle in most reports of clinical trials, often with medically important adverse consequences. Subgroup analyses as currently practiced are a source of many importantly wrong clinical decisions and regulatory decisions. There are occasions on which they produce medically useful conclusions, but there are, at present, far more occasions on which they do the opposite.”*
- Lots of scope for clever statistics to inform decision making
 - Exciting for many - ‘Intractable’
- Lots of difficult decisions to make!
 - Exciting for me – ‘Important’

- Legislation requires that marketing authorisation for a medicinal product shall be refused if:
 - (a) the risk-benefit balance is not considered to be favourable; or
 - (b) its therapeutic efficacy is insufficiently substantiated by the applicant; or
 - (c) its qualitative and quantitative composition is not as declared.

- Need to **establish** favourable risk-benefit across the breadth of the patient population?
 - Yes
- Need to **prove** efficacy in every subgroup?
 - No
- Are regulatory standards evolving toward this?
 - No

Examples of experience with clever statistics to address subgroups



Examples of difficult decisions
Vectibix - metastatic colorectal cancer

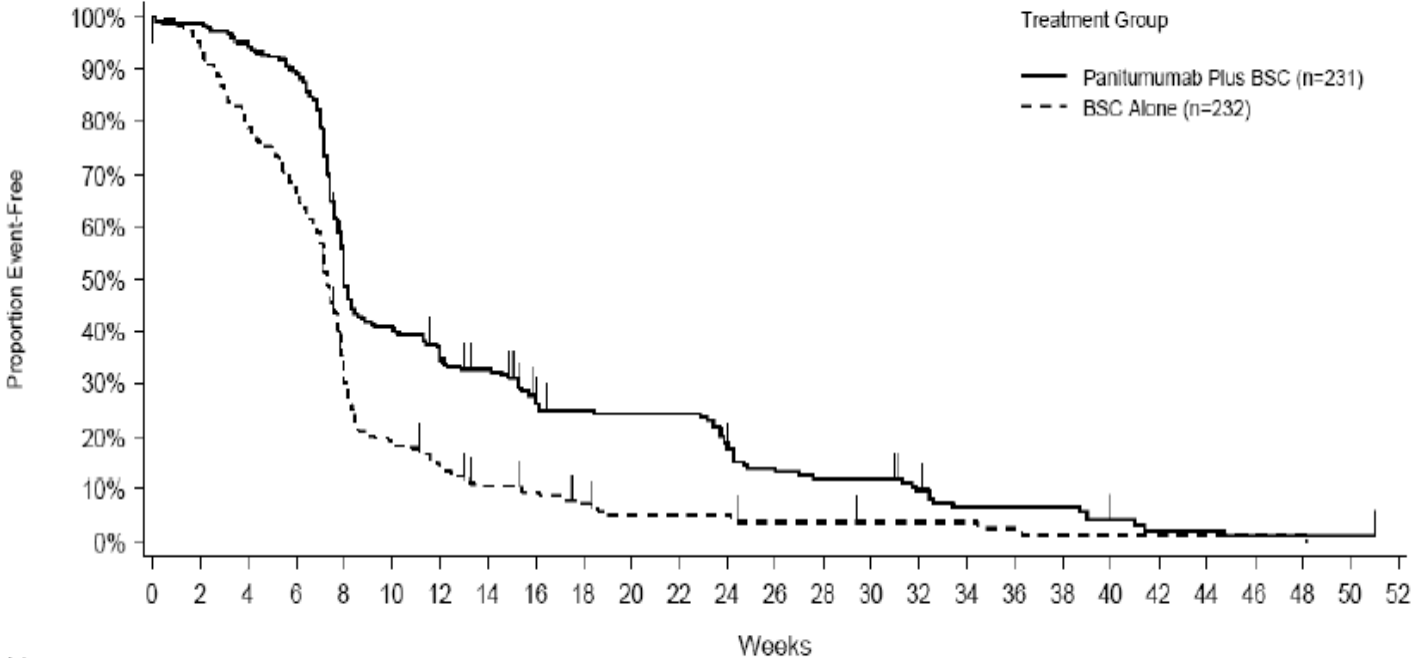
Study 20020408 was conducted as a pivotal clinical study in metastatic CRC with the following title:

“An open-label, randomised, phase 3 clinical trials of ABX-EGF plus best supportive care vs. best supportive care in patients with metastatic colorectal cancer”. Patients with progression in the best supportive care (BSC) alone-arm were allowed to cross over to receive panitumumab in an open-label extension study (Study 2003194).

Examples of difficult decisions

Vectibix - metastatic colorectal cancer

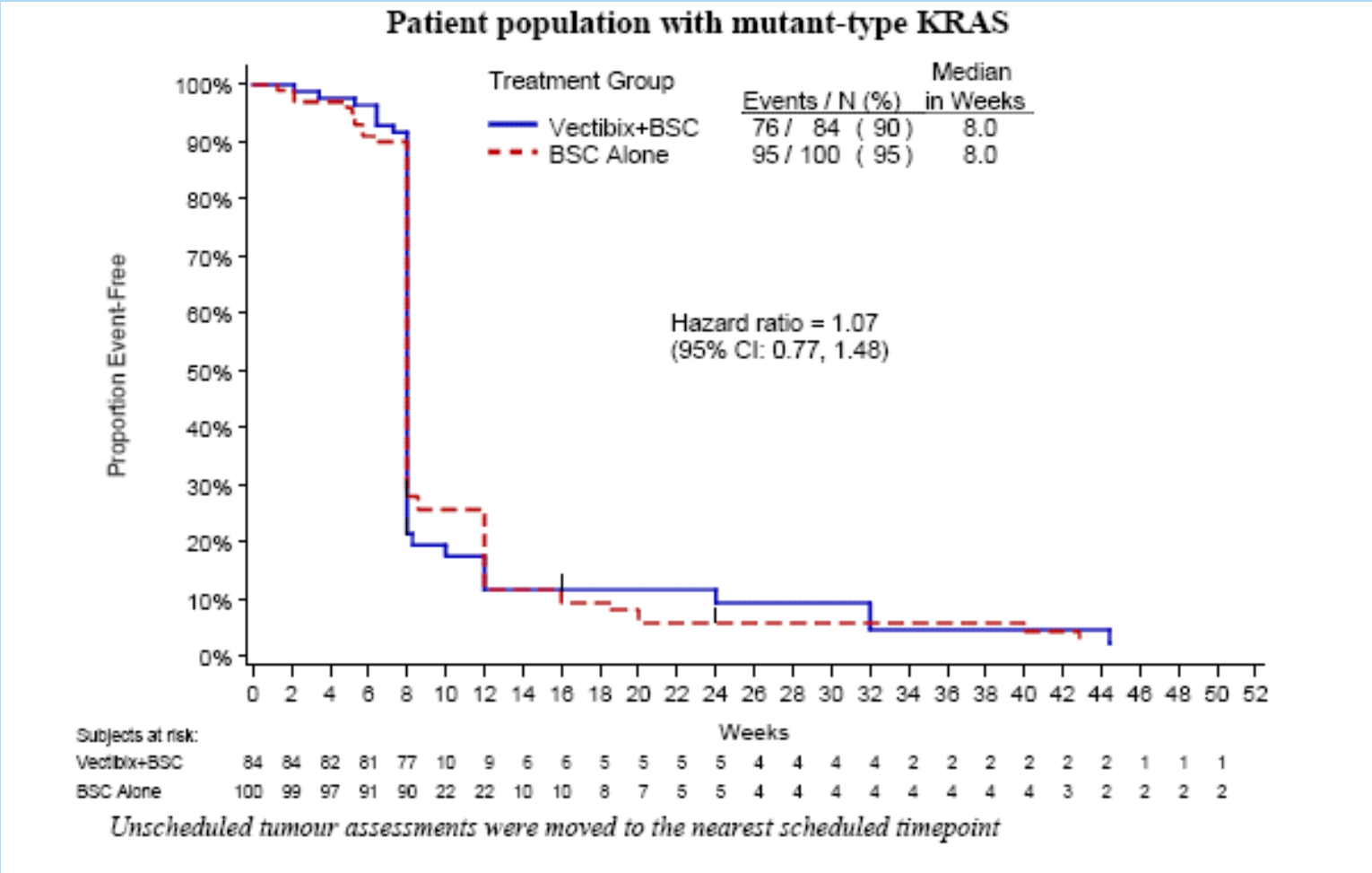
Figure 7. Study 20020408: Kaplan–Meier plot of PFS (ITT, IRC assessment)



Subjects at risk:

Panitumumab Plus BSC	231	217	209	197	118	85	76	65	49	41	40	31	22	19	19	13	8	8	8	5	2	2	1	1	1
BSC Alone	232	209	175	149	75	41	31	20	17	11	7	7	4	4	3	3	3	2	1	1	1	1	1	1	0

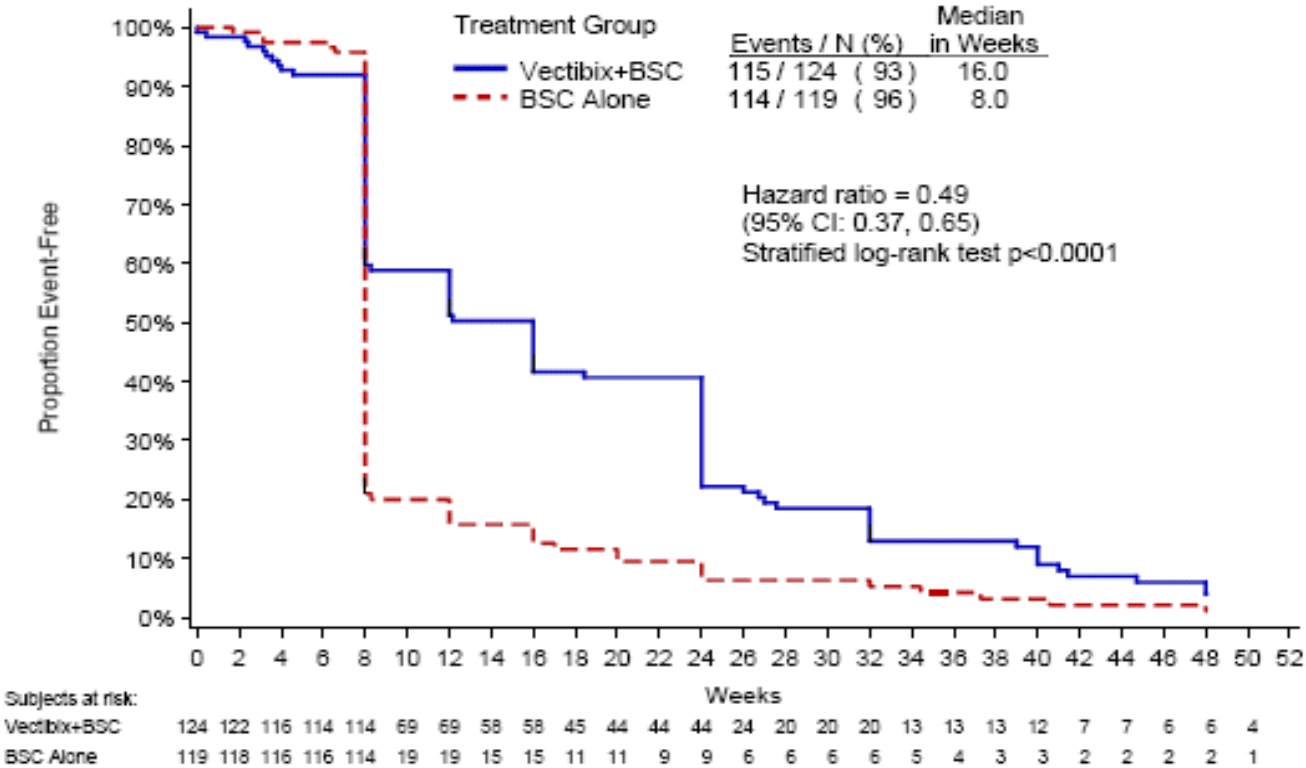
Vectibix mutant-type KRAS



Vectibix wild-type KRAS

Figure 16. Study 20020408 – Kaplan–Meier plot of PFS (ITT, time adjusted, IRC assessment)

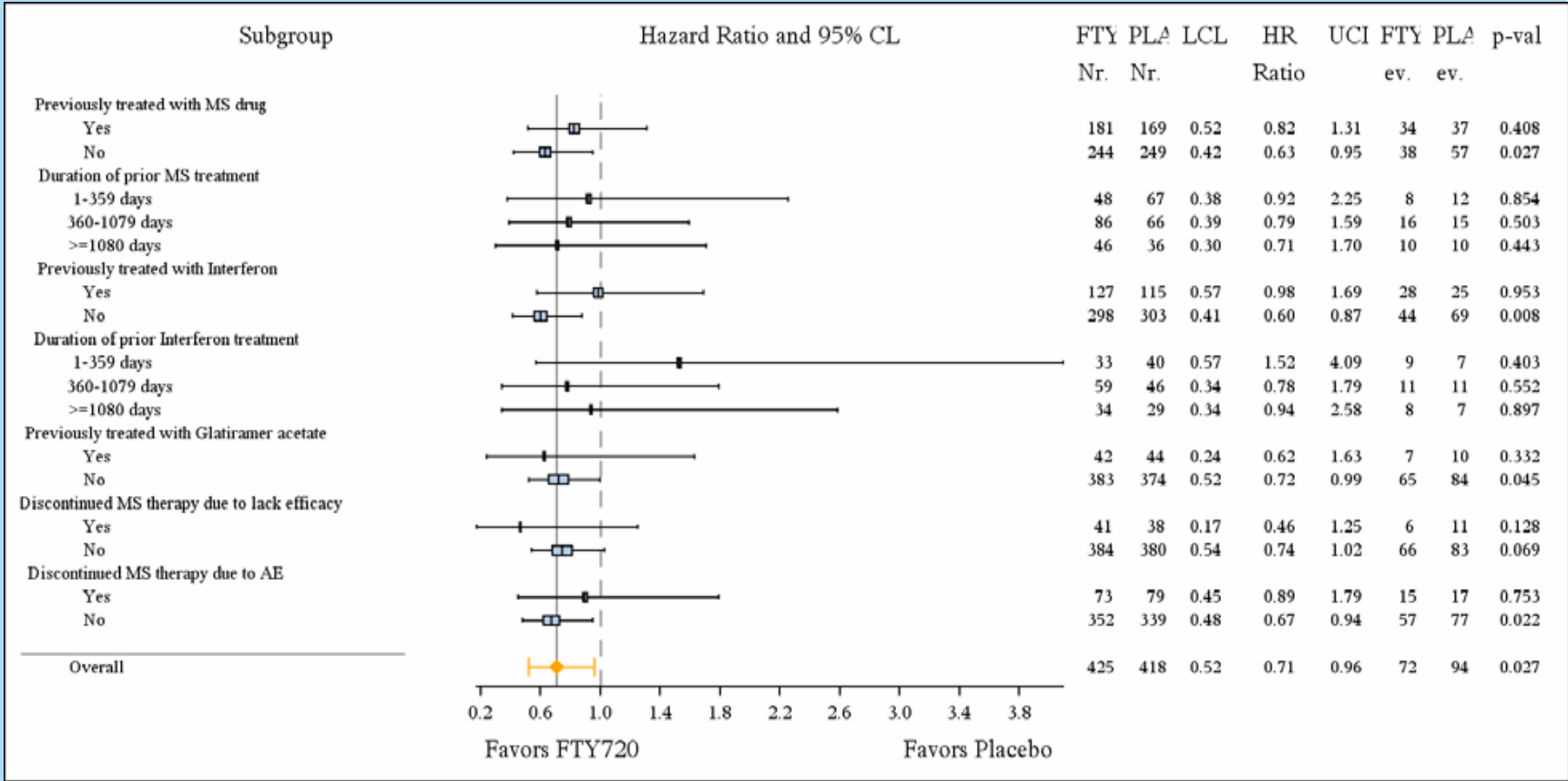
Patient population with wild-type KRAS



Unscheduled tumour assessments were moved to the nearest scheduled timepoint

Example Forest plot

How to interpret?



Problem statement

- If all patient 'groups' benefit to same extent (same Δ) then will see smaller estimated effects, even -ve effects for some groups, **by chance**
 - Action leads to False -ve decision
- If some patient 'groups' benefit to different extent (Δ_1 , Δ_2 , Δ_3 , possibly 0 or harm) then even more likely to see smaller estimated effects for some groups, **because it's real!**
 - To ignore leads to false +ve decision
- Cannot **know** which is true

Assumption is the mother of all foul ups...

- Assertion: Patients recruited to a clinical trial are not homogenous, nor should response to treatment be routinely assumed to be homogenous.
- Established regulatory thinking - see ICH E5 (intrinsic and extrinsic factors), GL on multiplicity
- Why such a problem? Statistics is all about not knowing the truth (=quantifying uncertainty)
 - For 'trial success' we pre-specify (n=1 or 2) and power
 - For exploration we face multiplicity (n=20-30) and little information

- Exploratory → **SUBGROUP G/L**
 - Assessment of 'internal consistency'
 - To assess strength of evidence and magnitude of efficacy across the population
 - Negative conclusions from subgroups because of
 - Efficacy
 - Benefit-Risk
 - Positive conclusions from subgroup analyses, in a 'failed' trial.
- The science of clinical trial methods and the art of decision making

Confirmatory vs Exploratory

- Confirmatory
 - Pre-planned confirmatory subgroup analysis
 - **MULTIPLICITY G/L**
 - Interesting science, less 'art'!
 - Some very promising use in drug development programmes (esp. oncology), though others ignore!

Negative conclusions from subgroups because of benefit-risk Gilenya (fingolimod) in relapsing-remitting multiple sclerosis

Annualized relapse rate (ratio: 0.40 and 0.46 for 1.25 mg and 0.5 mg, respectively).

Risk of disability progression, (0.5 mg: HR= 0.68, 95% CI: 0.50, 0.93, $p=0.017$; 1.25 mg: 0.5 mg: HR= 0.70, 95% CI: 0.52, 0.96, $p=0.024$) as compared to placebo.

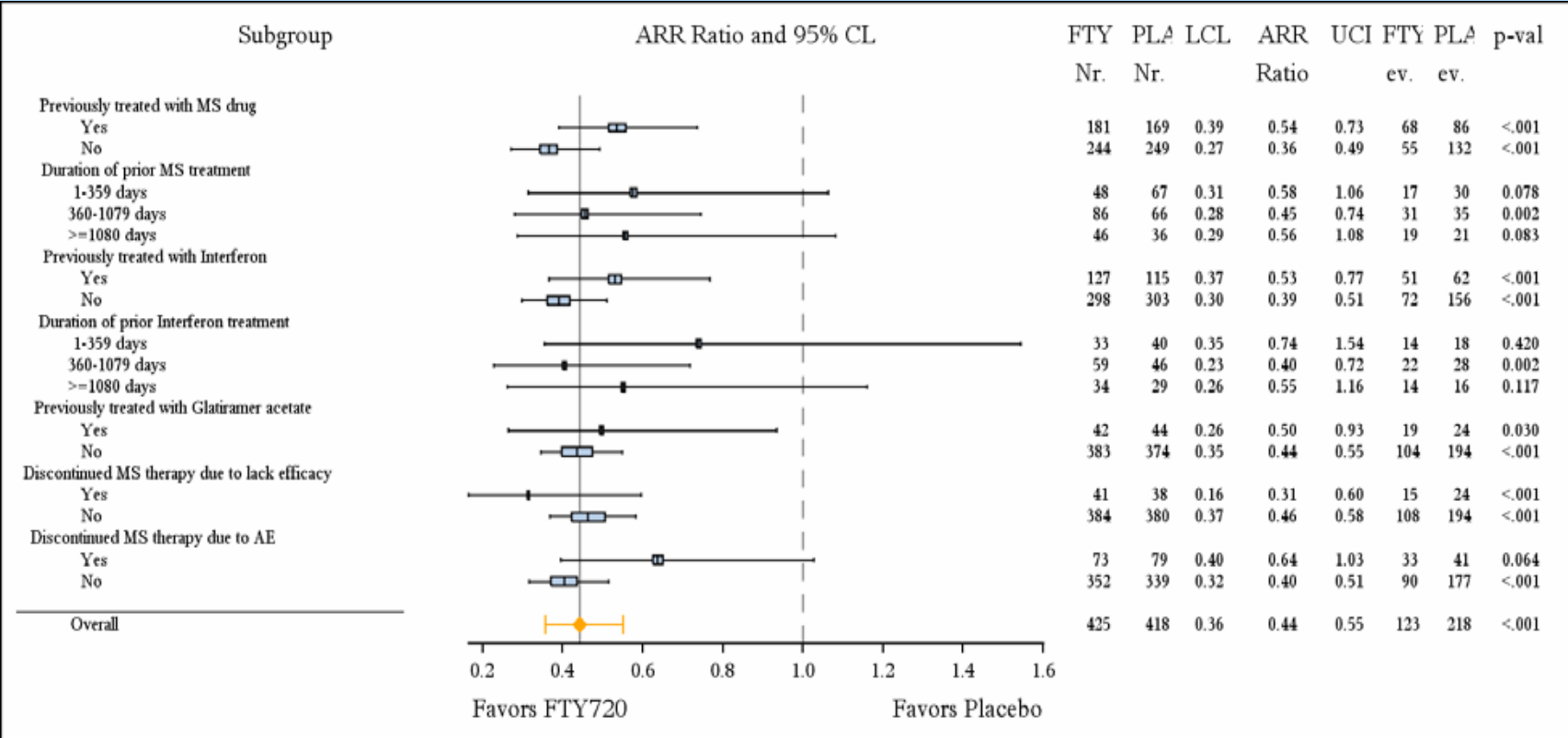
ARR (0.5 mg : 0.16, 95% CI: , 0.122, 0.212, $p<0.001$; 1.25 mg: 0.20, 95% CI: 0.157,0.264, $p<0.001$) versus interferon beta-1a (0.33, 95% CI: 0.262, 0.417), representing relative reductions of 52% and 38%, respectively. Relapse-free at month 12 was higher in the fingolimod treatment groups (79.8% for fingolimod 1.25 mg and 82.6% for fingolimod 0.5 mg treatment groups) compared to the interferon beta-1a i.m. group (69.3%) and this difference was statistically significant ($p<0.001$)

Negative conclusions from subgroups because of benefit-risk Gilenya (fingolimod) in relapsing-remitting multiple sclerosis

- Considering the heterogeneous safety profile of fingolimod, the benefit risk of fingolimod was considered negative by the CHMP in the indication initially applied for: *“Disease-modifying therapy in adults for the treatment of patients with relapsing multiple sclerosis to reduce the frequency of relapses and to delay the progression of disability”*.
- But – clearly effective – oral formulation – wish to look for favourable benefit-risk
 - Higher Unmet Medical Need
 - Bigger scope for benefit (in absolute terms)

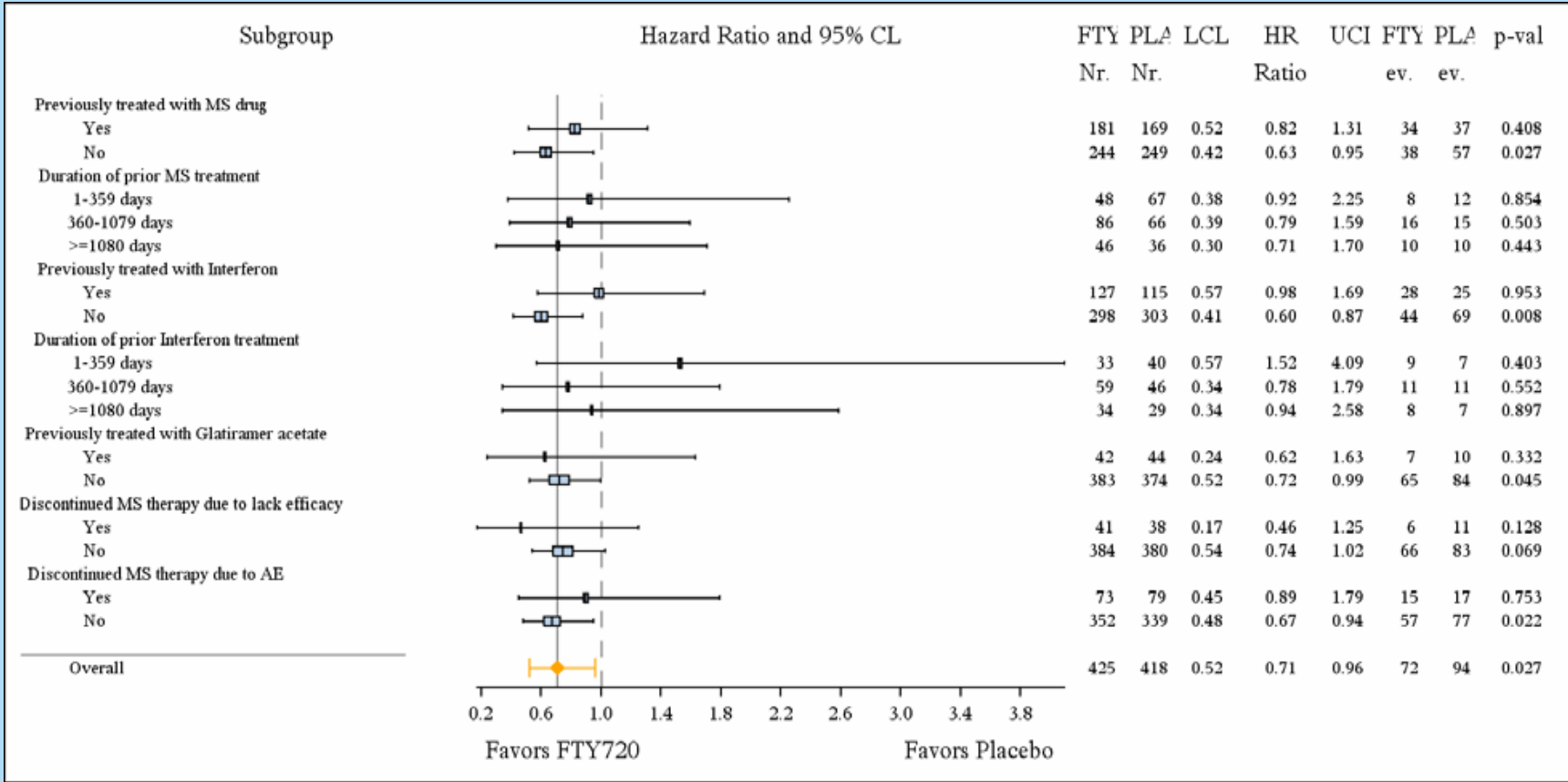


Forest plot of ARR in study D2301 comparing FTY720 0.5 mg vs. placebo, previous MS treatment (ITT population)



ARR subgroup = 0.37 (0.24,0.57), p<0.001

Forest plot of 3-month confirmed disability progression in study D2301 comparing FTY720 0.5 mg vs. placebo, previous MS treatment (ITT population)



Disability progression - subgroup = 0.78 (0.36, 1.68), p=0.521

Negative conclusions from subgroups because of benefit-risk
Gilenya (fingolimod) in relapsing-remitting multiple sclerosis

“The effects demonstrated versus placebo as well as the active comparator Avonex were considered clinically relevant in the overall population. Oral administration was considered to be of particular benefit given that the currently available therapies are using the parenteral route. A number of important safety concerns have been identified related to the mechanism of action that is first in the class. The safety of fingolimod (cardiac, ocular, immune, hepatic, and pulmonary systems) is characterised by a heterogeneous profile which required to recommend a restricted use in multiple sclerosis population as well as a number of measures to ensure safe and effective use of the product. Consistent treatment effects were observed in highly active group of RRMS patients as compared to the overall population and therefore this restricted population was recommended for the indication.”^{Slide 20}



Current practice

- Planning: Consider stratification of randomisation. Otherwise ignore. Keep fingers crossed
- Analysis: Standard? Thoughtless?
- Inference: Dismiss unfortunate findings

- This difficult problem deserves more attention. More attention should lead to more good decisions!

Some relevant questions for planning

- Is the patient population homogenous for progression of disease?
 - Prognostic factors (e.g. disease severity at baseline, failed previous treatments etc.)
- Is the patient population homogenous for (expected) response to treatment**?
 - Predictive factors (e.g. KRAS, 'tired' receptors)
 - **It's a matter of scale!
- What other potential sources of important heterogeneity?
 - E.g. Con-meds, region
- Categories are not mutually exclusive

If important heterogeneity might exist...

PtC on Adjustment for Baseline Covariates:

Section I.1: “If the effect of treatment is expected to vary substantially across important pre-specified subgroups (for example, age groups or race), then stratifying for these subgroups can help in interpreting the treatment effect and its consistency across these subgroups. This can also enhance the credibility of some subgroup analyses that are *a priori* of high interest. **If such an interaction is expected then the trial should be powered to the treatment effect within specific subgroups**”.

Also Section III.4:

Subgroup categorisation



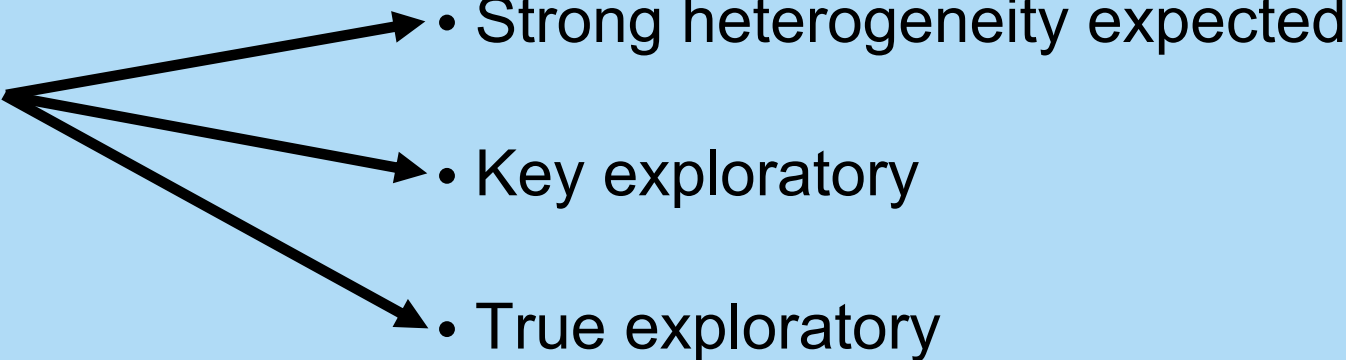
- Confirmatory
- Exploratory

Subgroup categorisation

- Confirmatory

- Confirmatory

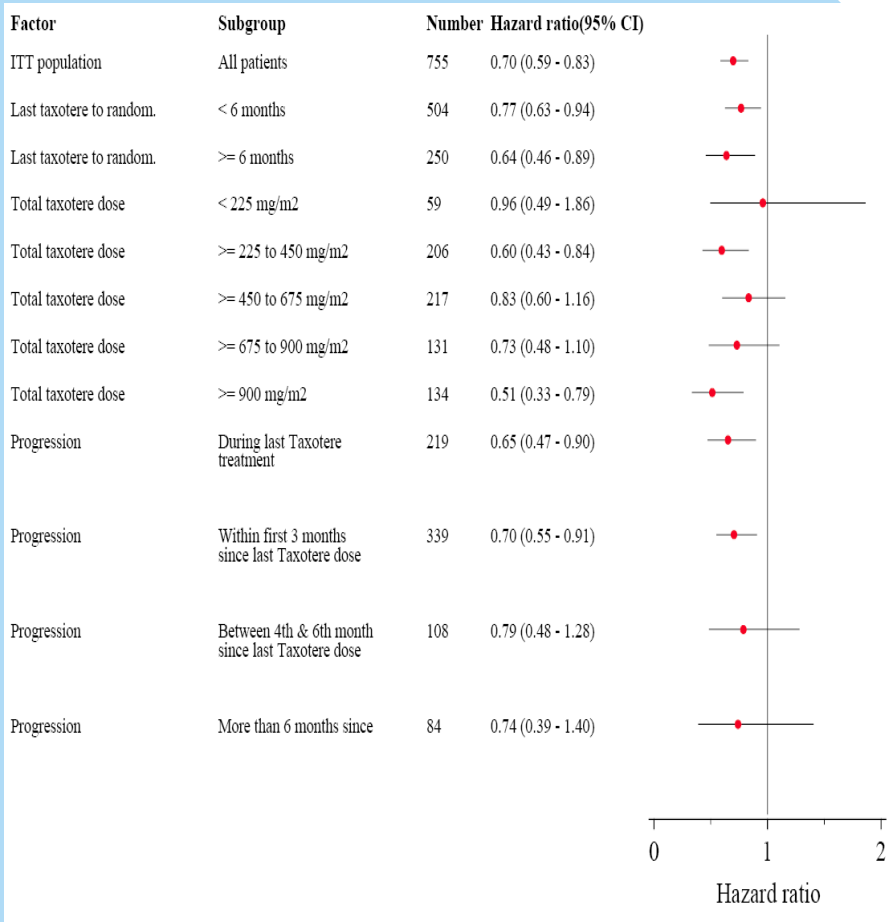
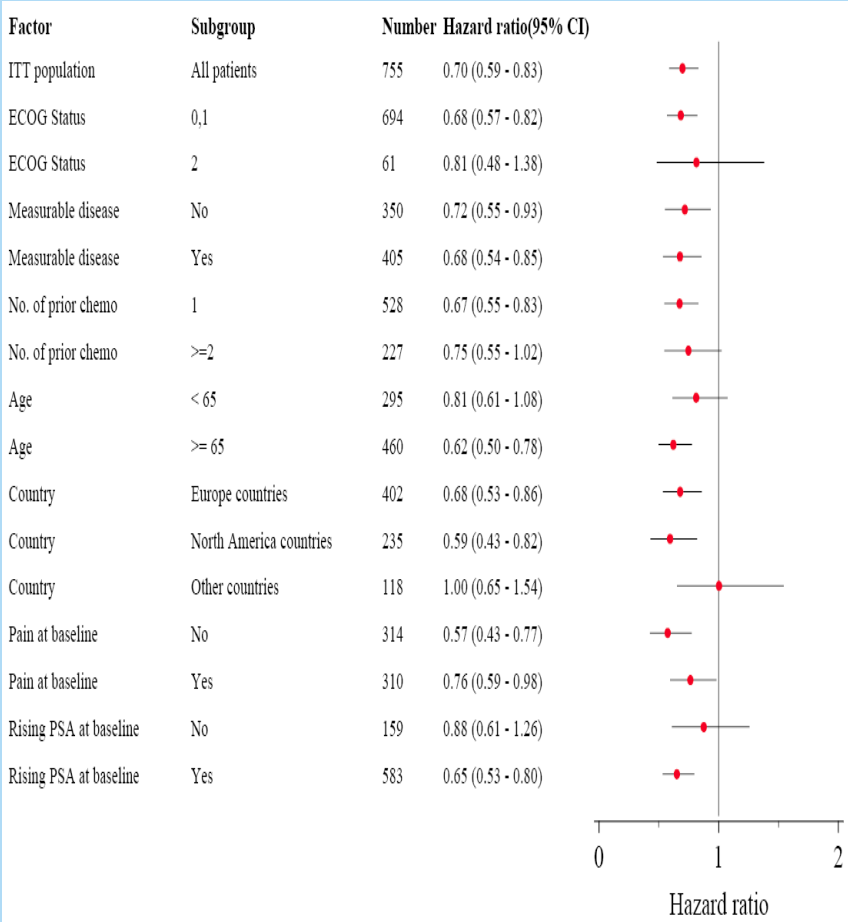
- Exploratory



Different types of subgroups

- Strong heterogeneity expected
 - Vectibix KRAS, Gender in Irritable Bowel Syndrome
 - Plan separate studies, or fully powered for subgroup analyses
- Key exploratory
 - Plausibility for some heterogeneity of response to treatment
- True exploratory
 - No rationale for heterogeneity
 - ASA in CV prevention / star sign

To optimise planning and interpretation of Forest plots?



To optimise planning and interpretation of Forest plots?

- Perform all subgroup analyses as ‘exploratory’ and try to discuss
- Perform zero subgroup analyses?
 - Can’t make any false negative decisions!
 - (Don’t) seek and you will (not) find!
 - Unacceptable, scientifically and regulatory
- Pre-specify subgroups of interest?
 - “This can also enhance the credibility of some subgroup analyses that are *a priori* of high interest. ”

To optimise planning and interpretation of Forest plots?

A proposal:

- Discuss prognostic factors within the patient population and predictive factors for treatment response in the protocol
- Identify key exploratory analyses (n=6-8?) based on:
 - Key demographic factors, including genomic factors, related to the mechanism of action / pharmacology would be highest priority (n=1-2?)
 - Key prognostic factors would be included. (n=2-4?)
 - Other factors that might plausibly be predictive for different response to treatment such as stage, severity or phenotype of disease, use of concomitant medications and possibly region or centre
- Would usually include stratification factors
- The fewer prognostic factors / predictive factors the fewer key exploratory analyses needed

To optimise planning and interpretation of Forest plots?

A proposal:

- **Do this properly and assessment can concentrate on these ‘key exploratory’ subgroups and ignore ‘true exploratory’ subgroups. Reduce, though not remove, data-driven discussions. Deal?**
- Perhaps “If the ‘targeted’ assessment of key subgroups demonstrates consistency of effect then further exploration of subgroups might not be necessary. If the ‘targeted’ assessment has been well planned, extreme data in other exploratory subgroups (but without a plausible link to treatment response) could generally be disregarded. If the discussion and pre-specification is incomplete then the assessor will by necessity need to take a more ad-hoc approach and will be forced to rely more on the observed data without the benefit of the structure given above that limits the number of subgroups examined.”

Sample size

- Even a focus on a few ‘key exploratory’ subgroups may result in misleading data and bad decisions.
- Agreed. So will ignoring subgroups.
- More information is good.
- Consider sample size / information in ‘key exploratory’ subgroups?

Problem statement
Analysis, Reporting and Decision making

- If all patient 'groups' benefit to same extent (same Δ) then will see smaller estimated effects, even -ve effects for some groups, **by chance**
 - Action leads to False -ve decision
- If some patient 'groups' benefit to different extent (Δ_1 , Δ_2 , Δ_3 , possibly 0 or harm) then even more likely to see smaller estimated effects for some groups, **because it's real!**
 - To ignore leads to false +ve decision
- Cannot **know** which is true
- No action is also a decision

References in available guidance

The MHRA logo is a dark blue oval with the letters "MHRA" in white, bold, uppercase font.

PtC on Multiplicity issues in clinical trials:

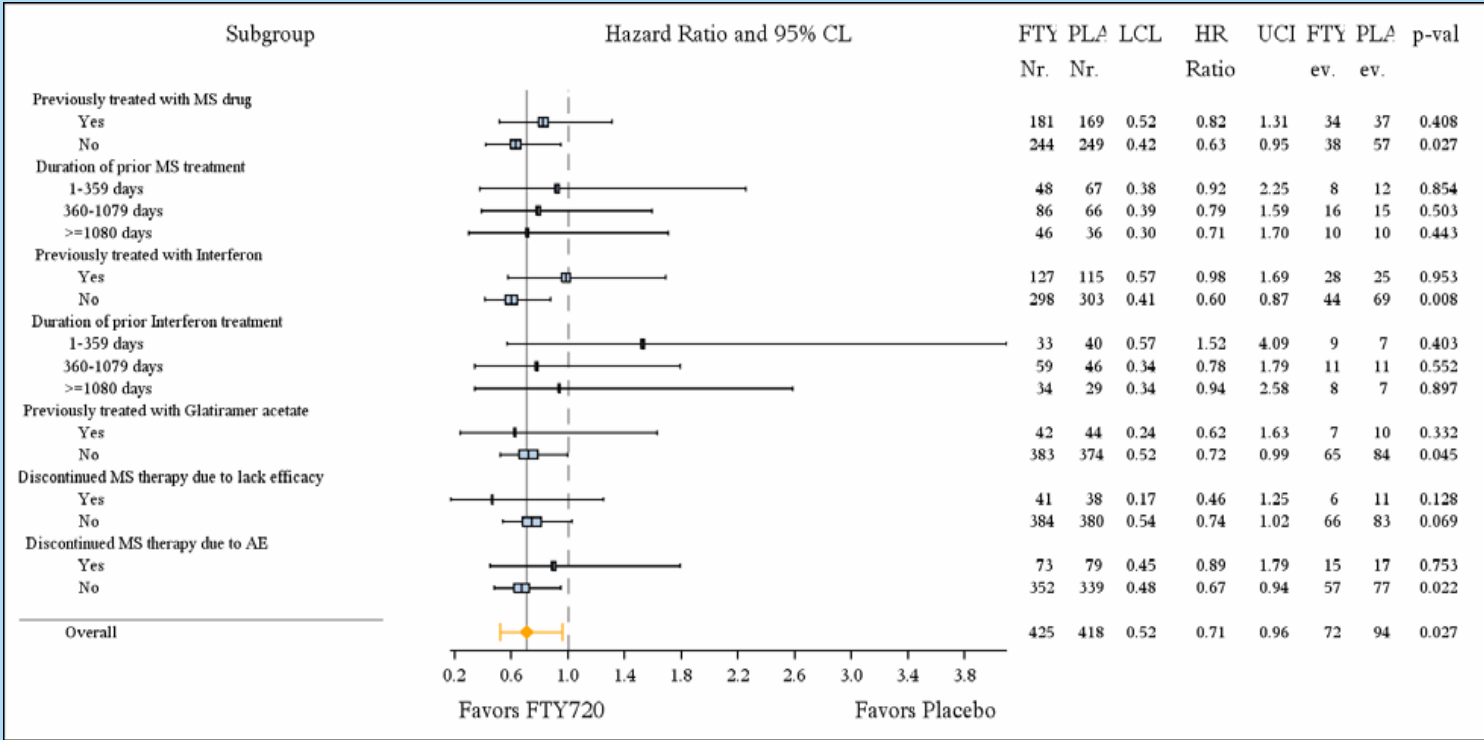
Sec. 4 Summary: “Reliable conclusions from subgroup analyses generally require pre-specification and appropriate statistical analysis strategies. **A license may be restricted if unexplained strong heterogeneity is found in important sub-populations, or if heterogeneity of the treatment effect can reasonably be assumed but cannot be sufficiently evaluated for important sub-populations**”.

Explanations that are too simplistic...

- *“the test for interaction is non-significant therefore ... ignore”*
 - *“we have performed many exploratory analyses therefore we will find some false negative results ... ignore”*
 - *“this analysis was not pre-specified therefore ... ignore”*
 - *“estimated effects in all subgroups are positive therefore ...ignore”*
- **All ‘consistent with’ efficacy BUT none alone can dismiss a valid concern**
- *“there is no biological plausibility to this result ... ignore”*
 - better, but prefer to discuss in advance – post-hoc justifications as difficult as post-hoc analyses!

Assessing 'consistency' – art or science?

- Try to develop a mathematical rule to **signal** groups that require further assessment?
 - NOT for decision making



Decision making

- Pharmacological rational / Biological plausibility
- Data / Strength of evidence / **Replication**



Decision making – rescuing a failed study?

- Pharmacological rational / Biological plausibility
- Data / Strength of evidence / **Replication**



Decision making – restricting an indication

- Pharmacological rational / Biological plausibility
- Data / Strength of evidence / **Replication**





Guideline

- 2013
- Long?
- No methods – just a principle, not discouragement to thoughtful (=novel) approaches.
- No increase to regulatory hurdles intended, but a reflection of current practice, consolidation of other guidelines and trying to describe strategies for assessment to improve planning, analysis, reporting and discussion in CTRs and in ARs!

Key messages

The MHRA logo is a dark blue oval with the letters 'MHRA' in white, bold, sans-serif font.

- The primary role of statistics when planning, analysing, reporting and drawing inferences from subgroup analyses is to properly inform decision making.
- Medicines don't only need to work, but need to have a favourable profile of benefits and risks in the licensed population.
- Complete homogeneity of response to treatment is rarely plausible
- Better planning ('key exploratory'?) → Better decision making. More up-front discussion and agreement with regulators?
- Bring all science to bear on decision making: greater weight to plausibility and replication than to p-values

Slide 40