Clinical trial design and HTA reimbursement

Professor Allan Hackshaw, Deputy Director

Cancer Research UK & UCL Cancer Trials Centre

To cover

Current challenges of trial design for reimbursement/health technology assessment

1. Surrogate markers
2. What to use as the control arm?
3. Is new drug better than standard, or has similar efficacy?
4. The value (or not) of patient reported outcomes
5. Can real-world evidence/data help?

German IQWiG General Methods 2015

Risk of bias at (i) study level and at (ii) outcome level

On the basis of these aspects, in randomized studies the risk of bias is summarized and classified as "high" or "low". A low risk of bias is present if it can be excluded with great probability that the results are relevantly biased.

The classification as "high" of the risk of bias of the result for an outcome does not lead to exclusion from the benefit assessment. This classification rather serves the discussion of heterogeneous study results and affects the certainty of the conclusion.

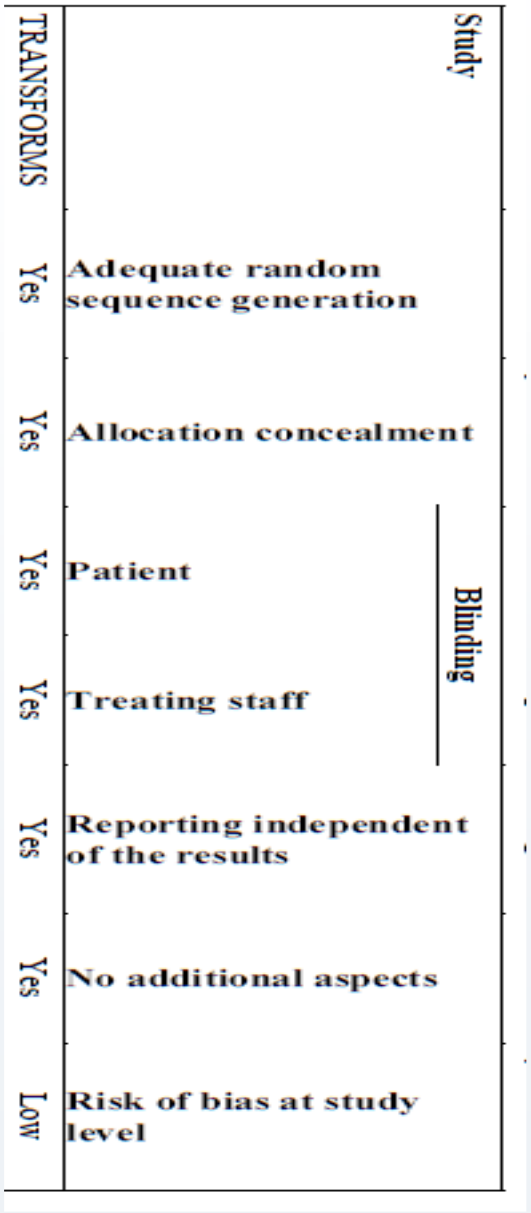
Bias is a big thing for HTAs. It can occur at:

(i) Design stage
(ii) Statistical analysis stage
(iii) Both of the above

Double-blind trial for multiple sclerosis

IQWiG General Methods 2015

Risk of bias at the study level



Study			TRANSFORMS
Adequate random sequence generation			Yes
Allocation concealment			Yes
Patient	Blinding		Yes
Treating staff	Blinding		Yes
Reporting independent of the results			Yes
No additional aspects			Yes
Risk of bias at study level			Low

Risk of bias for the outcome measures

Study		TRANSFORMS
Largely influenced by the potential for bias, validity & relevance of the measure, & appropriate analyses		
Study level		L
All-cause mortality	Outcomes	L
Relapses	Outcomes	L
Disability progression	Outcomes	L
Disability severity (MSFC-z)	Outcomes	H[a]
Fatigue (mFIS)	Outcomes	–[b]
Activities of daily living (PRIMUS activities)	Outcomes	–[b]
Health status (EQ-5D VAS)	Outcomes	H[c]
Health-related quality of life (PRIMUS QoL)	Outcomes	–[b]
SAEs	Outcomes	L
Discontinuation due to AEs	Outcomes	L
Infections	Outcomes	L
Flu-like illness	Outcomes	L
Constipation	Outcomes	L

L: low H: high

1. Surrogate markers
Problem with ‘hard’ endpoints like overall survival

Issues with, for example, overall survival/mortality now

1. Several treatments for a disorder: outcomes much better than years ago

2. Multiple lines of therapy after the trial treatments, especially when imbalanced between the trial groups

3. Unblinding patients at (i) end of trial or (ii) when they relapse: control patients cross over to the new therapy

Cannot distinguish effects of the trial treatments from the other therapies

Requirements for benefit assessment in Germany and England – overview and comparison

	Benefit assessment in Germany (G-BA/IQWIG)	Single Technology Appraisal in England (NICE)
Surrogate endpoints	Validation study applicable to the disease, its severity, the intervention, and the comparator required (exception: very serious diseases).	Accepted if correlation with final endpoint is strong or outcome measures are large.

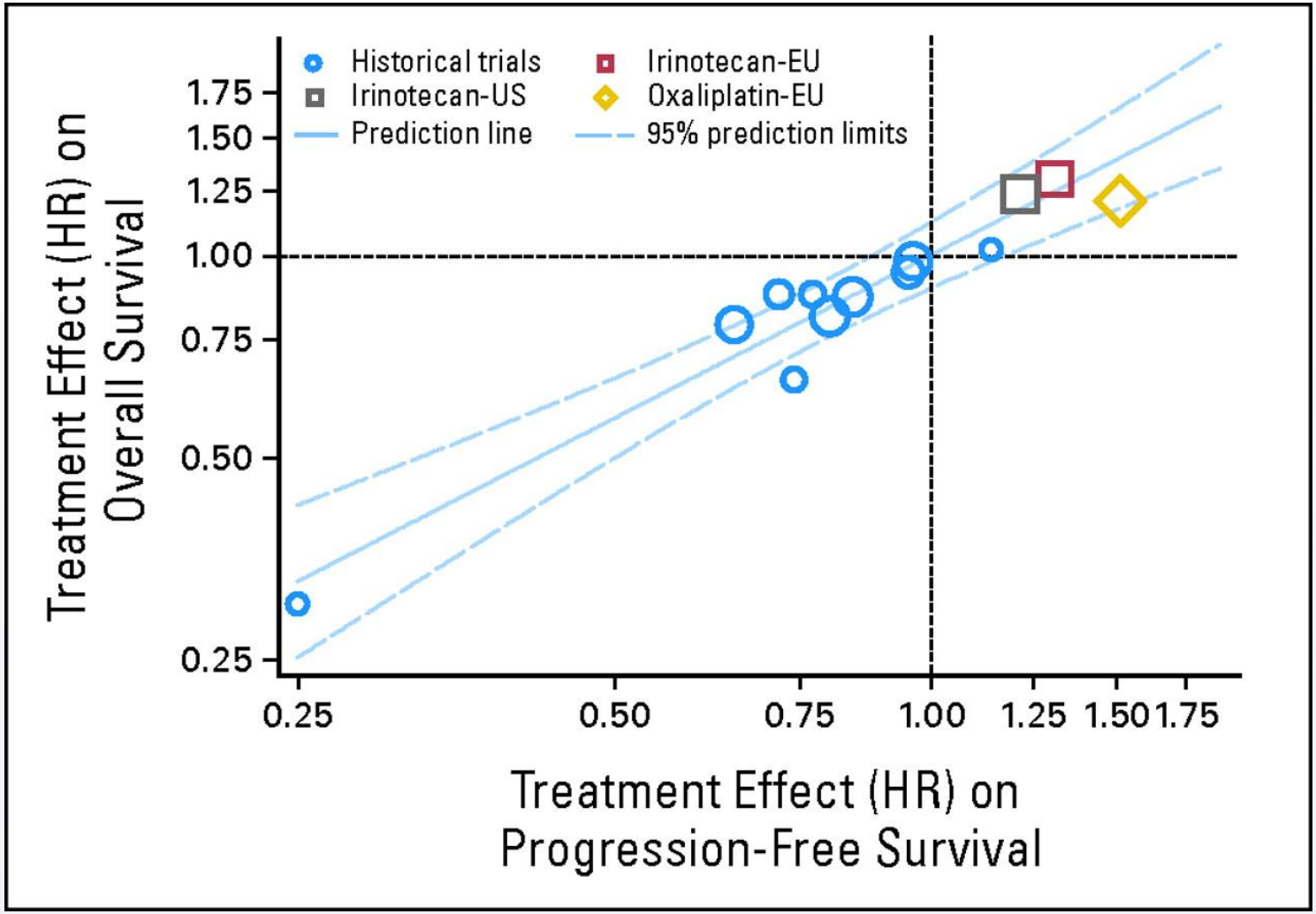
Variability in how different EU HTAs consider surrogates

NICE HTA more likely to accept surrogates than the German G-BA/IQWiG, but not easy
Can make trial design difficult in order to please several agencies

Ivandic *Health Economics Review* 2014, 4:12

International Journal of Technology Assessment in Health Care, 30:3 (2014), 312–324.

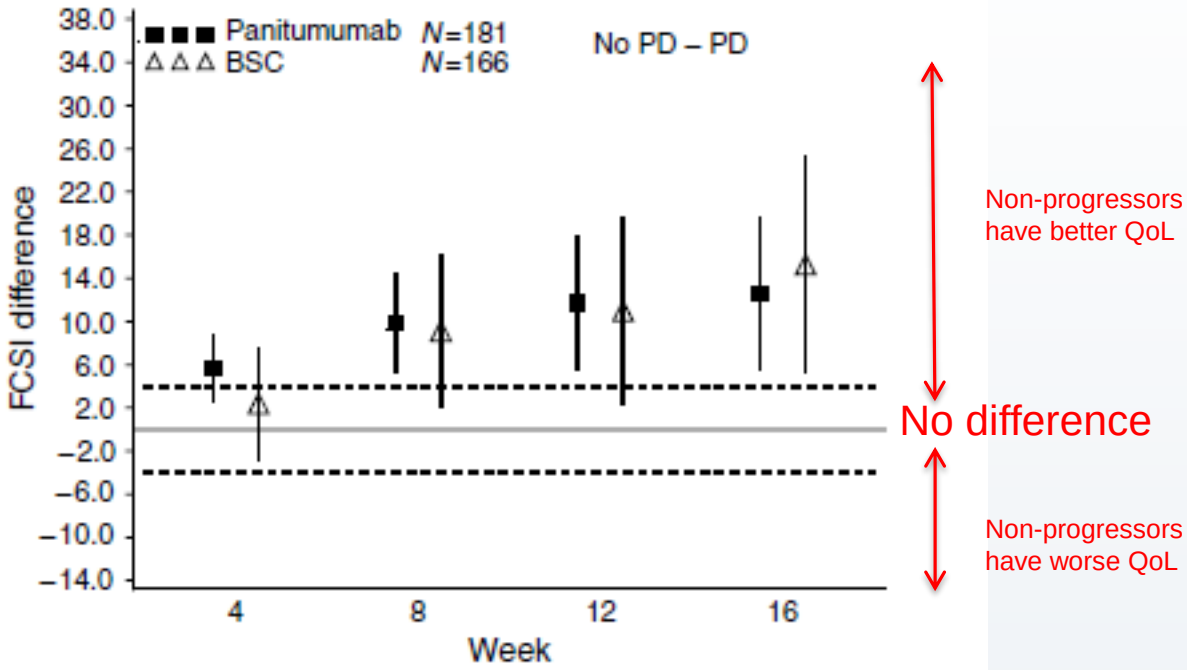
Evidence for PFS as surrogate for OS in advanced colorectal cancer (13 trials)
Correlation coefficient = 0.99
Very high correlation indicates that PFS is a good surrogate for OS
But you need several trials to show this!



Buyse et al. JCO 2007;25:5218-5224

HTAs are focussed on patient relevant endpoints
Large phase III trial of panitumumab for metastatic colorectal cancer

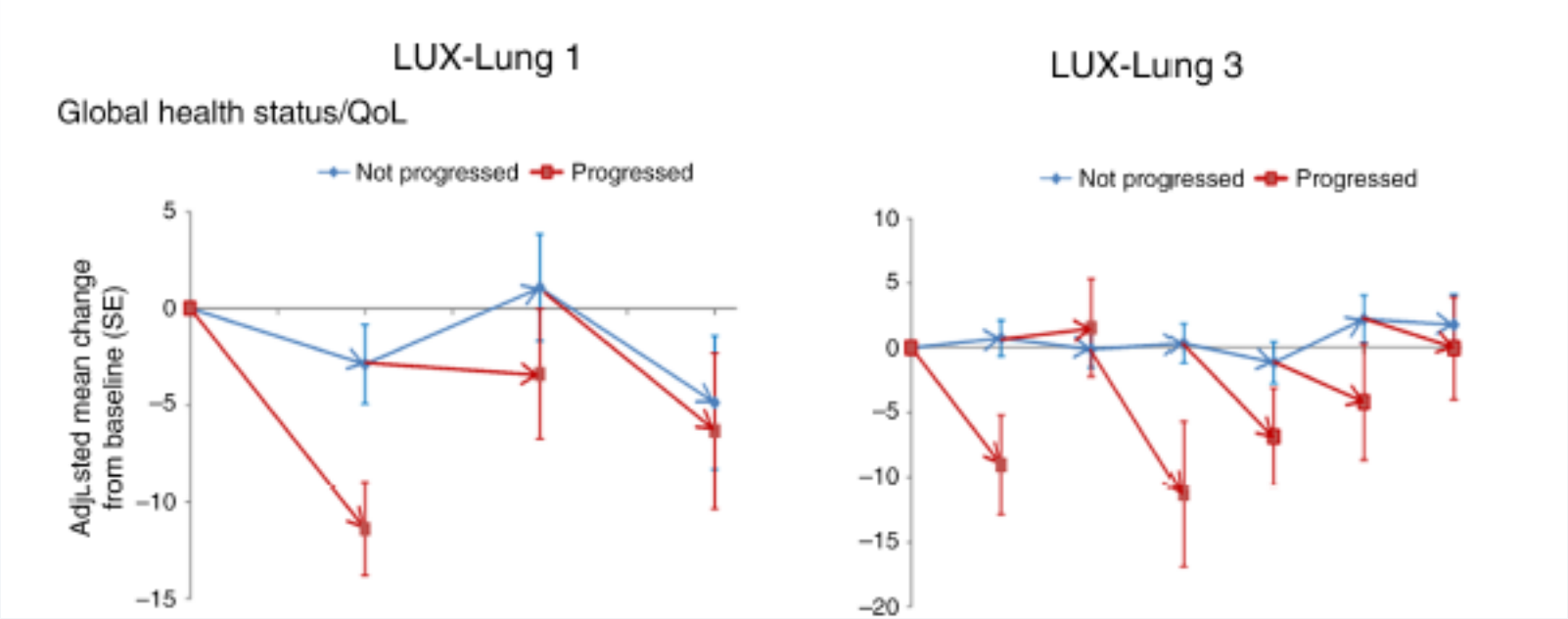
Colorectal-specific QoL (symptoms)



Note: dashed line at –3.94 represents a minimum clinical important difference (MCID) anchored to a ½ point change in ECOG score (Mathias *et al*, 2006). Difference is for no PD – PD.

British Journal of Cancer (2007) 97, 1469 – 1474

2 phase III trials of afatinib for advanced lung cancer

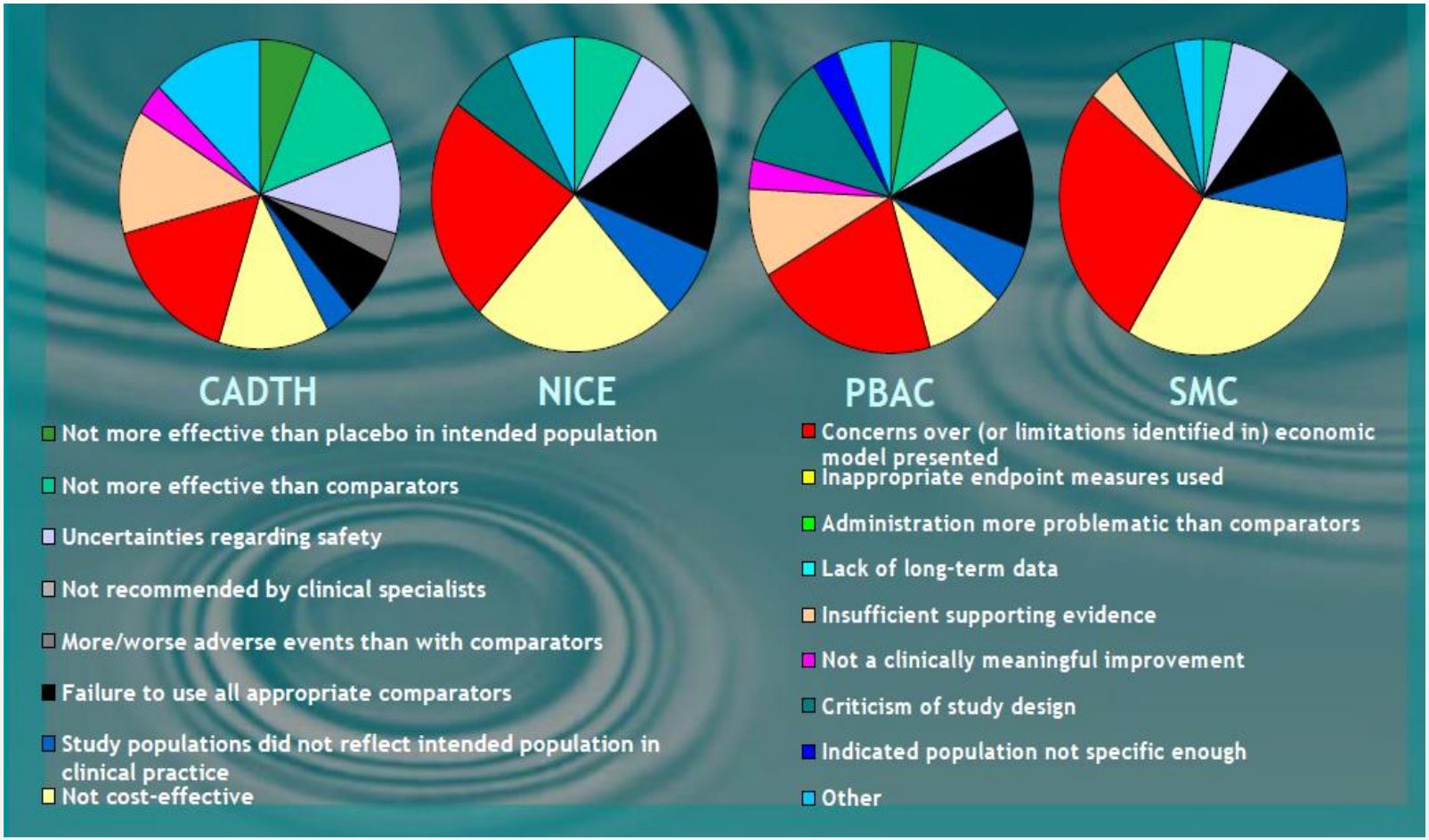


QoL gets worse when patients progress.
Delaying progression seems to maintain QoL/not make it worse

This should help support the value of surrogate endpoints in HTA submissions

2. Choosing comparators/controls

Reasons for rejection by 4 HTAs (Canada, England, Australia, Scotland)
Look at the size of the black sections: inappropriate comparators



Cann et al. Heron evidence development

2005-2008

Example: Fingolimod (Gilyena) for relapsing-remitting multiple sclerosis: licenced by EMA

2 trials	Number of patients	Annualised relapse rates (%)	
FREEDOMS		Fingolimod	Placebo
All patients	1272	18	40
Population 1b	160	21	54
TRANSFORMS		Fingolimod	Avonex
All patients	1292	16	33
Population 1b	374	25	51

Very clear efficacy results in all patients and the 1b subgroup (largest subgroup)

Population 1b: highly active relapsing–remitting MS who have an unchanged or increased relapse rate or ongoing severe relapses compared with the previous year despite treatment with beta interferon

Health Technology Assessment (HTA) review

- England NICE declined it in 2011
- They accepted placebo for FREEDOMS trial, but not Avonex
- Avonex in patients who have previously not responded to it, could be substandard treatment (continued use of an ineffective therapy)
- Uncertain evidence of the efficacy of Avonex
- No direct comparison with another drug used in the UK (Tysabri/natalizumab)
- Avonex as the comparator could make fingolimod appear to have greater cost-effectiveness

HTA review

- UK NICE changed their mind and approved it in 2012, due to:
- Indirect comparisons on treatment efficacy between all drugs available
- Lower price
- Strong support from patients
- These factors helped not having the control group NICE preferred

- IQWiG did not accept placebo for the FREEDOM trial, but they did accept Avonex for TRANSFORMS

Challenge for sponsors:

(a) have 1 trial with different control arms relevant to each major geographical region: more efficient and cheaper

(b) have separate trials in each region: more likely to show statistically significant results for each comparator type

Multi-arm trials…..

Comparison of ixekizumab with etanercept or placebo in moderate-to-severe psoriasis (UNCOVER-2 and UNCOVER-3): results from two phase 3 randomised trials

Lancet 2015; 386: 541–51

4 trial arms:

- Ixekizumab every 2 weeks
- Ixekizumab every 4 weeks
- Etanercept
- Placebo

Multi-arm trials: evidence satisfies:

- Comparison with standard therapy (head to head)
- Use of placebo (for subjective endpoints)
- Powered for both comparisons above
- 2 independent trials, giving same answer

Very clear results
Approved by IQWiG & NICE

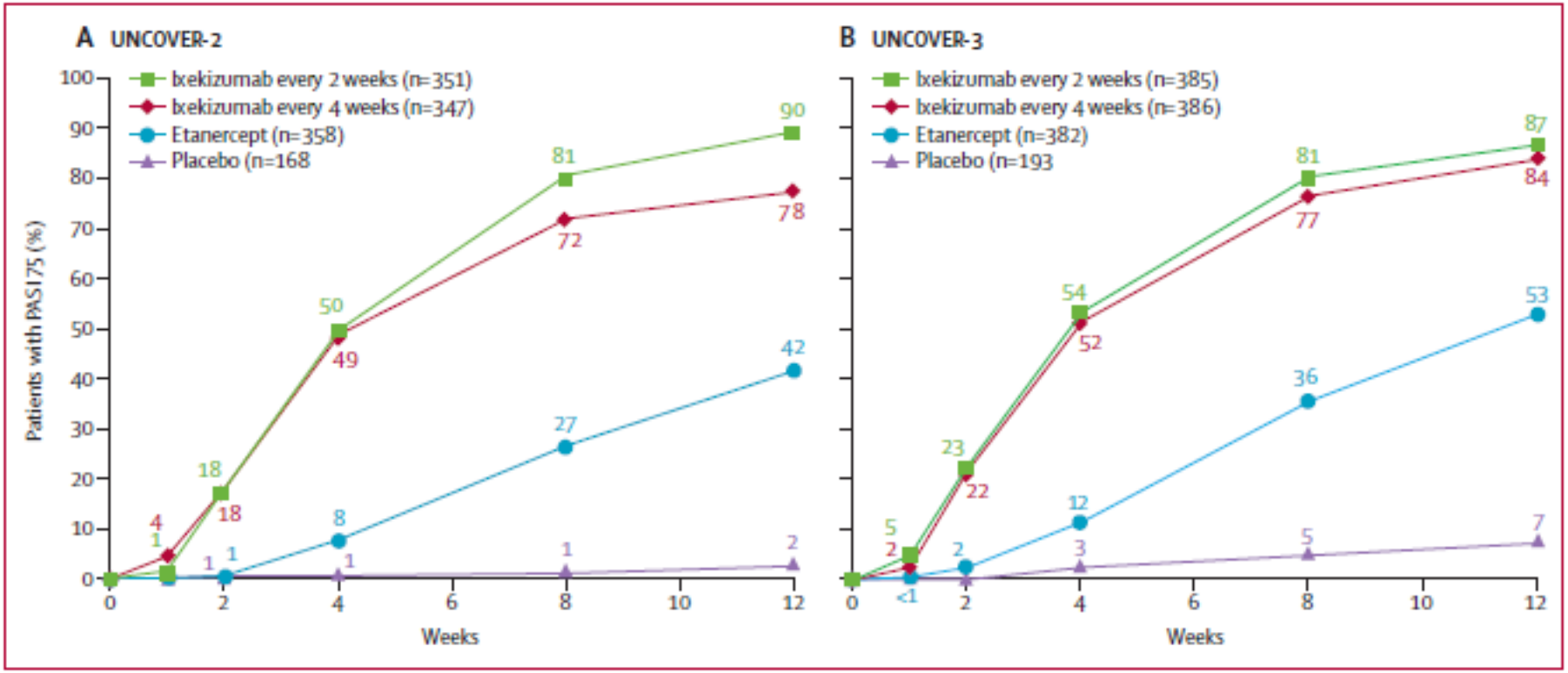


Figure 2: Proportion of patients achieving PASI75 from baseline through to week 12 in UNCOVER-2 (A) and UNCOVER-3 (B)

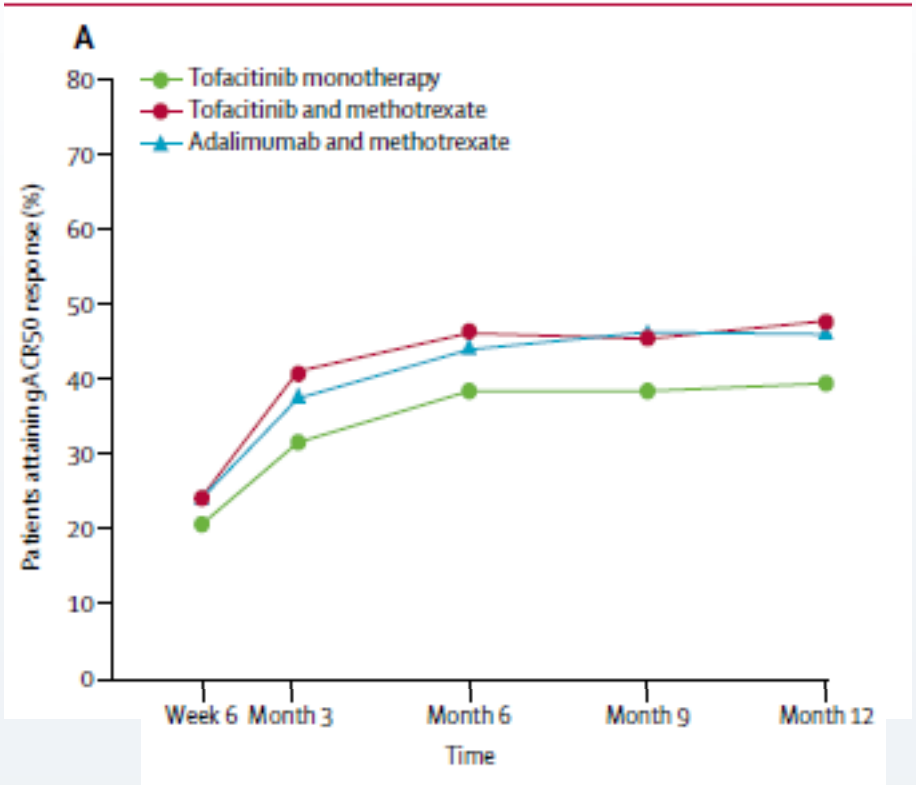
3. Non-inferiority

- If standard therapies already have high success, then difficult to get better therapy
- Find new treatments that have a similar effect to the standard, but the new one is:
  - Cheaper
  - Easier to administer (eg oral drug instead of injection)
  - Fewer side effects
  - Better quality of life/PRO

Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy):

Fleischmann et al Lancet 2017

Primary endpoint: ACR50
Validated and well-established
Covers both patient reported outcomes and clinician assessment



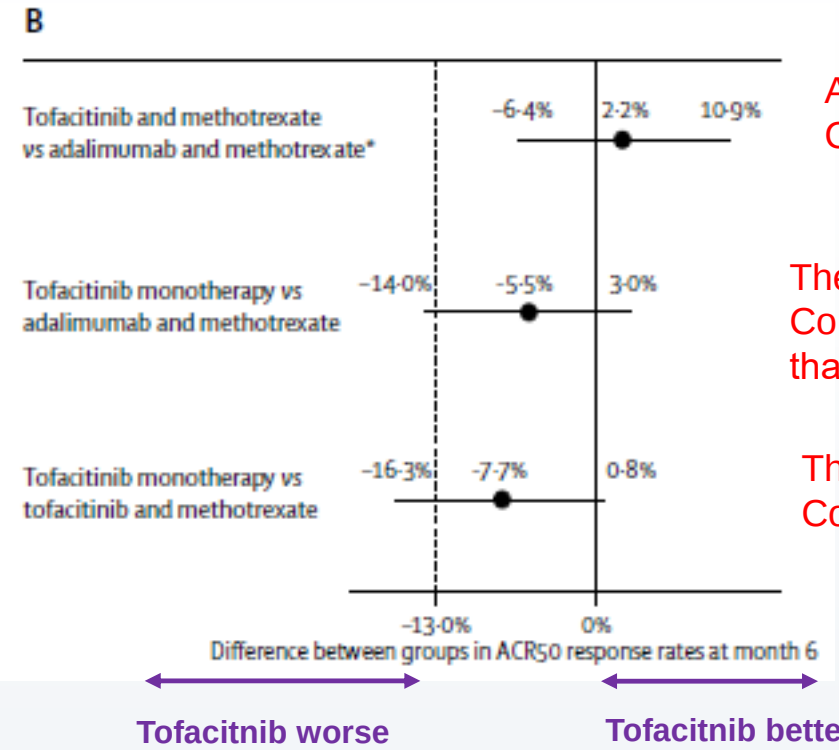
Patients in all groups improved up to 6 months
After this time, they remained steady (no more responders)

Allowable difference in the design

- Expected ACR50 is 35% using Adalimumab
- What loss in ACR50 is considered OK? They chose 13%

- They said this is half the efficacy for combination therapy vs placebo (which ~26%).
- This is one way of specifying the allowable difference
- But ultimately do you agree with it?

Interpreting non-inferiority



All of the 95%CI does not overlap the 13% margin.
Conclude: Tof/Meth just as good as Adal/Meth

The CI just overlaps the 13% margin.
Conclude: it is possible that the true effect of Tof is worse than Adal/Meth. But only just……

The CI clearly overlaps the 13% margin.
Conclude: Tof could be worse than Tof/Meth.

We don't use standard p-values (we use confidence intervals, CI)

CI indicates what the true effect of a treatment is likely to be

Need to consider:

- How important is the disorder?
- What if the person ‘fails’, i.e. salvage (does this outweigh the non-efficacy benefits of the therapy?)

- Small allowable differences require very large trials
- Large allowable differences require small trials – but then not accepted by HTAs

Example: assisted reproduction (delivery of progesterone)

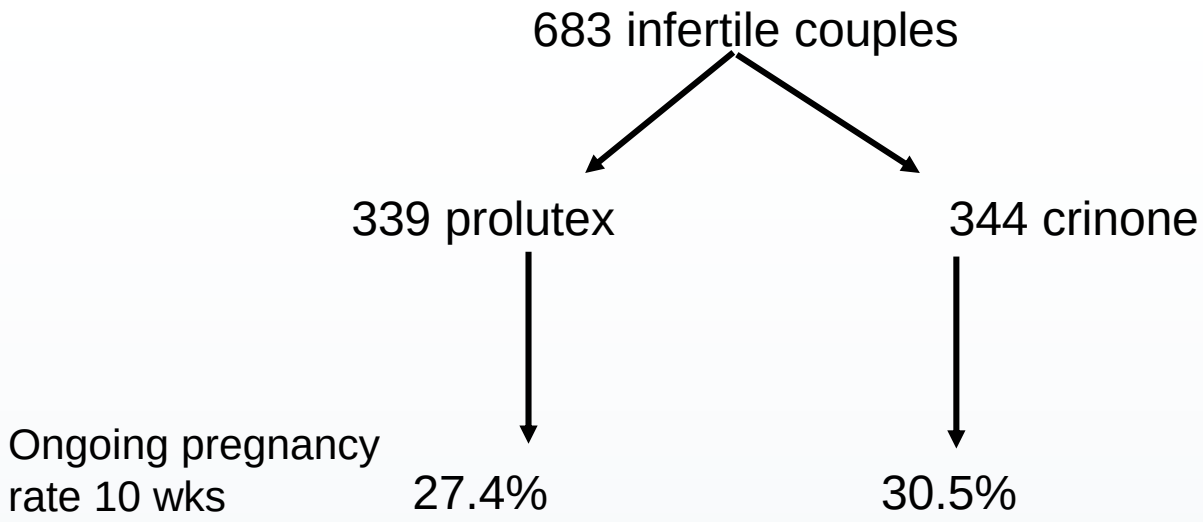
Prolutex (subcutaneous injection) versus
Crinone, standard (vaginal preparation): discomfort when applying, other side effects

Primary outcome: viable pregnancy at 10 weeks
~30% for crinone

Allowable difference: How much lower than this would you accept for Prolutex - 'you' being (i) a health professional or (ii) infertile couple?

Researchers chose 20% (i.e. allowable difference of 10%)

Lockwood et al. Fertil Steril 2014



Risk difference = -3.1%, 95% CI –9.9 to +3.7

95% CI does not go below -10%: conclude prolutex is non-inferior to crinone

Best estimate of effect for prolutex: pregnancy rate lower by 3.1% (ie 3 fewer among 100 treated)

But the true effect could be as high as nearly 10 fewer among 100 treated

4. Patient-reported outcomes (PRO)

- Many treatments become safer, and more people survive previously fatal disorders, there is much value in how patients ‘feel/function’ when taking a new therapy:
  - Daily living/working and well-being
  - Physical functioning
  - Mental functioning
  - Social functioning
  - Personal costs
  - Satisfaction with treatment
  - Adherence/compliance to treatment
- PROs are especially good for chronic conditions

- HTAs usually like PROs: patient-relevant outcomes

Open label randomised trial of secukinumab vs fumaric acid esters for plaque psoriasis (IQWiG review): chronic disorder

Dermatology Life Quality Index (validated questionnaire): range is 0 to 30
Low score is good (i.e. 0 or 1)

- 10 questions, covering: symptoms, embarrassment, shopping and home care,
- clothes, social and leisure, sport, work or study, close relationships, sex, treatment.

Is this a 'good' PRO measure?

- It covers several aspects of person's well-being and ability to function
- Impact on work: important for payers and employers (societal benefit)

Dermatology Life Quality Index: low score is good (i.e. 0 or 1)

Median time taken for patients to get to low score

Secukinumab	Fumaric acid ester
2.3 months	5.7 months

Hazard ratio: 4.49 ($p<0.001$)
Patients given secukinumab >4 times more likely to get to score of 0-1

Clear benefit of treatment on QoL

German IQWiG HTA assessment

Table 17: Extent of added benefit at outcome level: secukinumab vs. fumaric acid esters

Health-related quality of life		
DLQI (0 or 1)	Median: 2.33 vs. 5.68 months HR: 4.49 [2.69; 7.47] HR[c]: 0.22 [0.13; 0.37] $p < 0.001$ probability: "hint"	Outcome category: health-related quality of life $CI_u < 0.75$, risk $\geq 5\%$[e] added benefit, extent: "major"

Due to lack of blinding of patients

Due to big effect
They approved it

Double blind randomised trial of ixekizumab vs ustekinumab in patients with plaque psoriasis: same QoL measures as before

SF-36 physical functioning (PCS)
Mean change in score from baseline

Ixekizumab	Ustekinumab
5.5 units	3.5 units

Difference: 1.93
P-value=0.009

Statistically significant difference. But…….
PCS scale is 0 to 100……hence a difference of 1.93 is clinically very small

Conclusion on this endpoint by IQWiG

Lesser benefit/added benefit not proven[d]

This endpoint probably inappropriate/insensitive for this disorder and treatment

5. Can real world evidence (observational studies) help?

Survey of journal editors

- Preference for RCTs (due to design strengths)
- Lack of knowledge/training in RWE by peer-reviewers & editors
- Lack of transparency: RCTs are relatively easy to design/analyse & therefore review - unlike observational studies which often require complex analyses
- But they encourage more RWE submissions

International Journal of Technology Assessment in Health Care, 34:1 (2018), 111–119.

Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies

Amr Makady, MSc[1,2,], Renske ten Ham, MSc[2], Anthonius de Boer, MD[2], Hans Hillege, PhD[3], Olaf Klungel, PhD[2], Wim Goettsch, PhD[1,2], on behalf of GetReal Workpackage 1*

VALUE IN HEALTH 20 (2017) 520–532

Literature review of HTA guidance documents/publications
And interviews with representatives of the HTAs

6 agencies examined:
Sweden (TLV)
England (NICE)
Germany (IQWiG)
France (HAS)
Italy (AIFA)
Netherlands (ZIN)

Variability in how they handle RWE

They have hierarchy of evidence: RWE/observational studies almost always placed lower than RCTs

Real world evidence (RWE): observational studies

Strengths	Limitations
More generalizable patient & clinicians	Lots of variability between patients, healthcare systems
Can find uncommon side effects	Difficult to attribute side effects to the new therapy, or due to something else
Few/no eligibility criteria	Often cannot check source data (unless significant resources)
Patients follow routine practice (pragmatic)	No comparable control group (confounding): difficult to attribute differences in endpoints to the intervention
No need for randomisation, and distribution of different therapies	Known/unknown major reasons why some patients take the new therapy, and others do not (bias)
Might be useful as the comparator for single arm trials	Difficult to reliably estimate treatment effects & safety profile (could find no effect, or over-/under-estimate effect)

RCT & observational studies when both done

Examples	Findings
Flu vaccine in the elderly	RCTs shows same benefit as observational studies (preventing flu)
Hormone replacement therapy in post-menopausal women	Observational: 50% lower risk of heart disease RCTs: no effect
Estrogen-alone HRT	Observational: 30% higher risk of breast cancer RCTs: no/smaller effect
High β-carotene	Observational: 31% lower risk of cardiovascular mortality RCTs: 12% higher risk
Statins & cancer patients	Observational: 20-40% lower risk cancer recurrence/death RCTs: no effect

Example: Dapagliflozin: diabetes control

- Clear beneficial effects from the randomised trials.

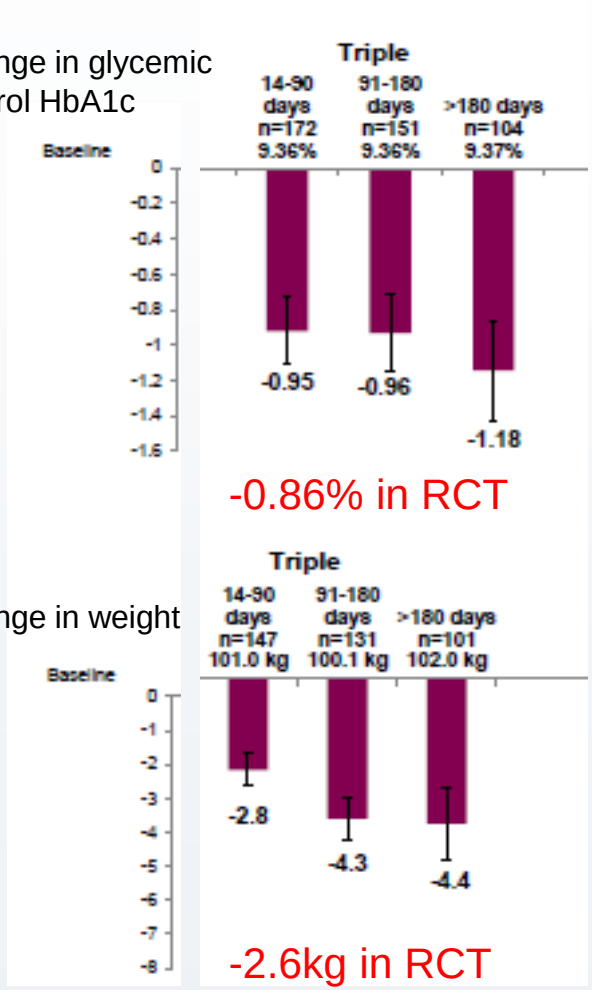
- Real-world evidence, included in NICE submission.
- National database: UK Clinical Practice Research Datalink (Nov 2012-Sep 2014), 684 primary care units

- Retrospective longitudinal observational study (12 million patients)

NB: large, national, representative of routine practice patients

Days on treatment

Change in glycemic control HbA1c



-0.86% in RCT

Change in weight

-2.6kg in RCT

RWE patients had slightly worse baseline glycemic control & higher body weight than the RCT patients

But: effect of dapagliflozin on these 2 endpoints generally similar to the RCT

Hence, RWE supports the RCT evidence

Accepted by England NICE

Closing remarks

- Differences in expectations of trial designs by regulatory agencies and health technology agencies
- Multi-arm trials likely to become more common to satisfy different organisations, where feasible
- Patient reported outcomes essential: but need to choose the right ones
- RWE might be OK, but methods need careful planning (to minimise bias and confounding)