Bayesian evidence synthesis – does it lead to more stringent criteria for benefit assessment?

Simon Wandel, Expert Statistical Methodologist | Cologne, 17 June 2016

NOVARTIS
ONCOLOGY

Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein.

Agenda

Background

P-values, Bayes factors and probabilities

A holistic view of evidence

Bayesian evidence synthesis: case study

Conclusion

A primer on Bayesian statistics (1/2)

- Probability is the key to Bayesian Statistics
- “Bayesian Statistics = Applied Probability Calculus”
- All uncertainties are expressed probabilistically
  - e.g. probability hazard ratio < 0.75 = 0.5
  - or probability hazard ratio > 1 = 0.28
  - makes the Bayesian approach inherently simple

A primer on Bayesian statistics (2/2)

- Key elements of Bayesian statistics
  - distribution of the data (model)

$$p(D|\theta)$$

  - prior distribution of the parameter

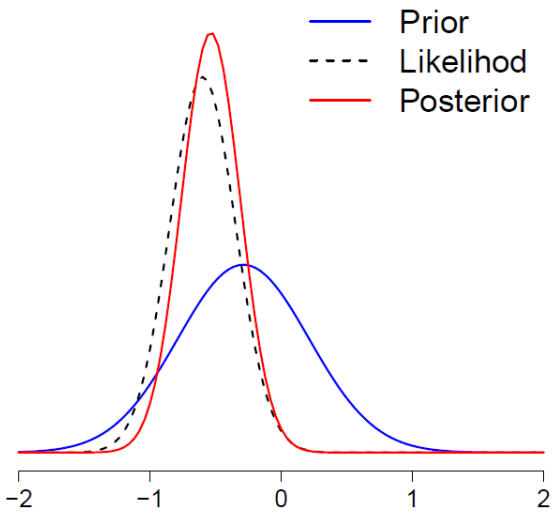
$$p(\theta)$$

  - posterior distribution of the parameter

$$p(\theta|D)$$

  - updating rule (Bayes Theorem)

$$p(\theta|D) = \frac{P(D|\theta)p(\theta)}{p(D)} \propto p(D|\theta)p(\theta)$$



NOVARTIS ONCOLOGY

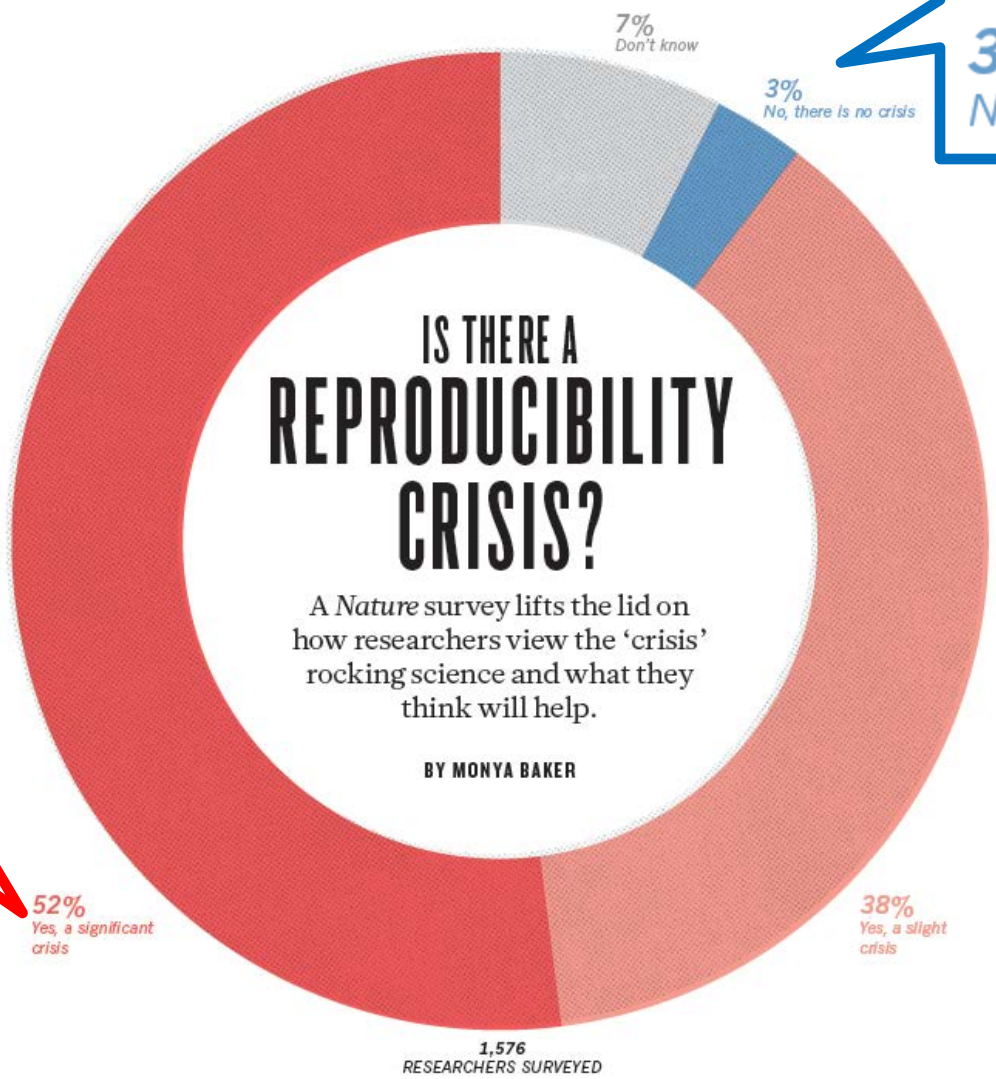
Revised standards for statistical evidence

- Johnson, VE. *Revised standards for statistical evidence.* (PNAS 2013).
  - relies on Johnson VE. *Uniformly most powerful Bayesian tests.* (Ann Stat. 2013)
  - main idea
    - use Bayes factor (BF) to quantify strength of evidence
    - for some special models uniformly most powerful Bayesian tests (use BF)

In terms of classical hypothesis tests, these evidence standards mandate the conduct of tests at the 0.005 or 0.001 level of significance.

- In what context has this statement been made?

The scientific reproducibility crisis



3%
No, there is no crisis

52%
Yes, a significant crisis

- Baker M. *Is there a reproducibility crisis?*

NOVARTIS ONCOLOGY

What’s the issue and what should we do about it?

Agenda

Background

P-values, Bayes factors and probabilities

A holistic view of evidence

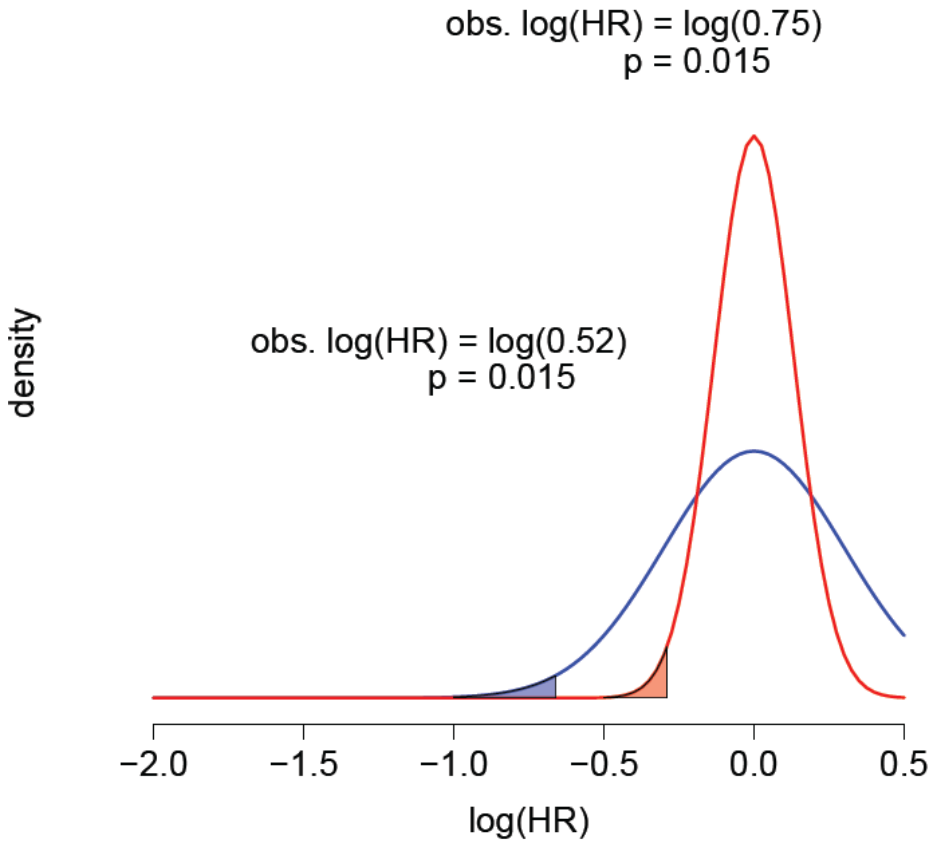
Bayesian evidence synthesis: case study

Conclusion

P-values

- P-values can be a source of great confusion (Wasserstein & Lazar 2016)
- I will use a simple example here
  - hazard ratio (HR) treatment vs. control, HR < 1 favors treatment
  - one-sided test at 2.5%
  - two studies (1:1 randomized) with 44 and 228 events
  - for the larger study: 90% power at $HR_A = 0.65$
  - presented are
    - p-values
    - Bayes factors
    - posterior probabilities

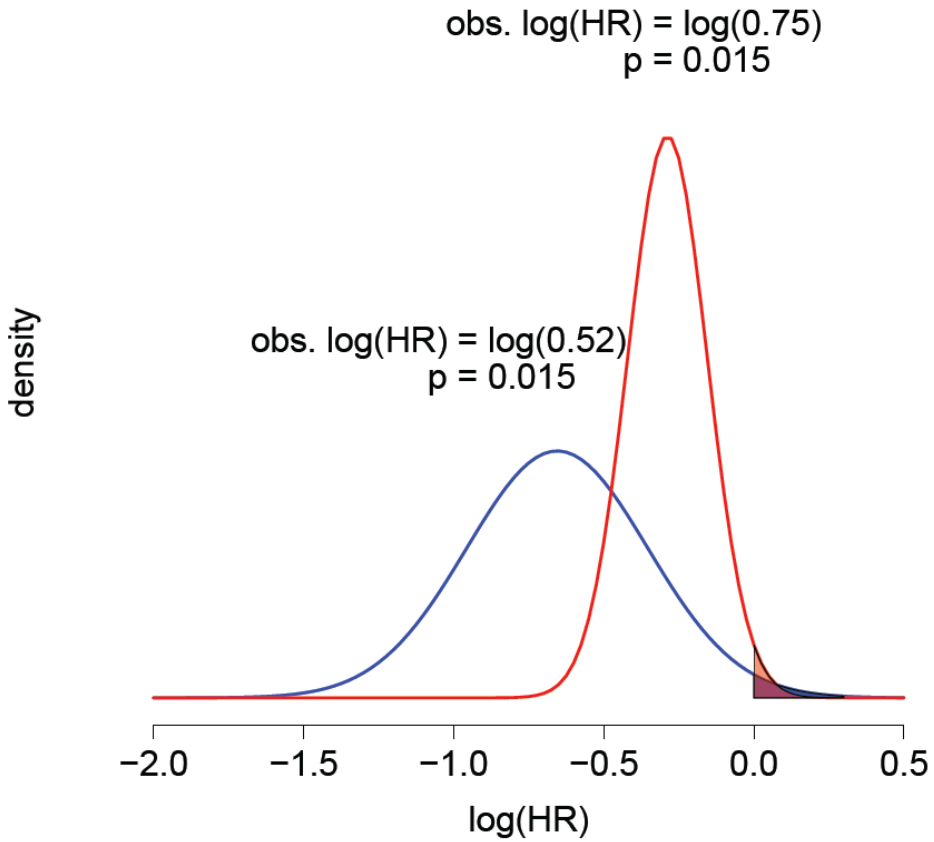
Example (1/4): p-values



- both studies show a significant effect (identical p-values)
- favor HR_A=0.65 or HR_0=1?

NOVARTIS ONCOLOGY

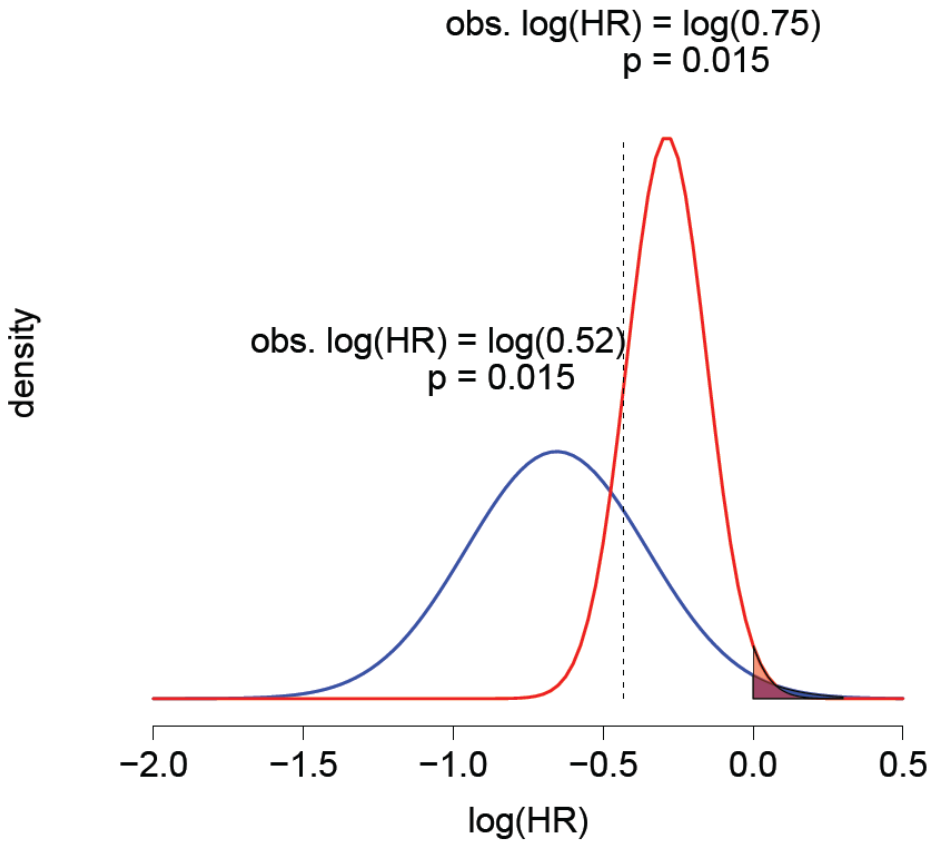
Example (2/4): posterior distribution



- both studies show a significant effect (identical p-values)
- favor HR_A=0.65 or HR_0=1?

NOVARTIS ONCOLOGY

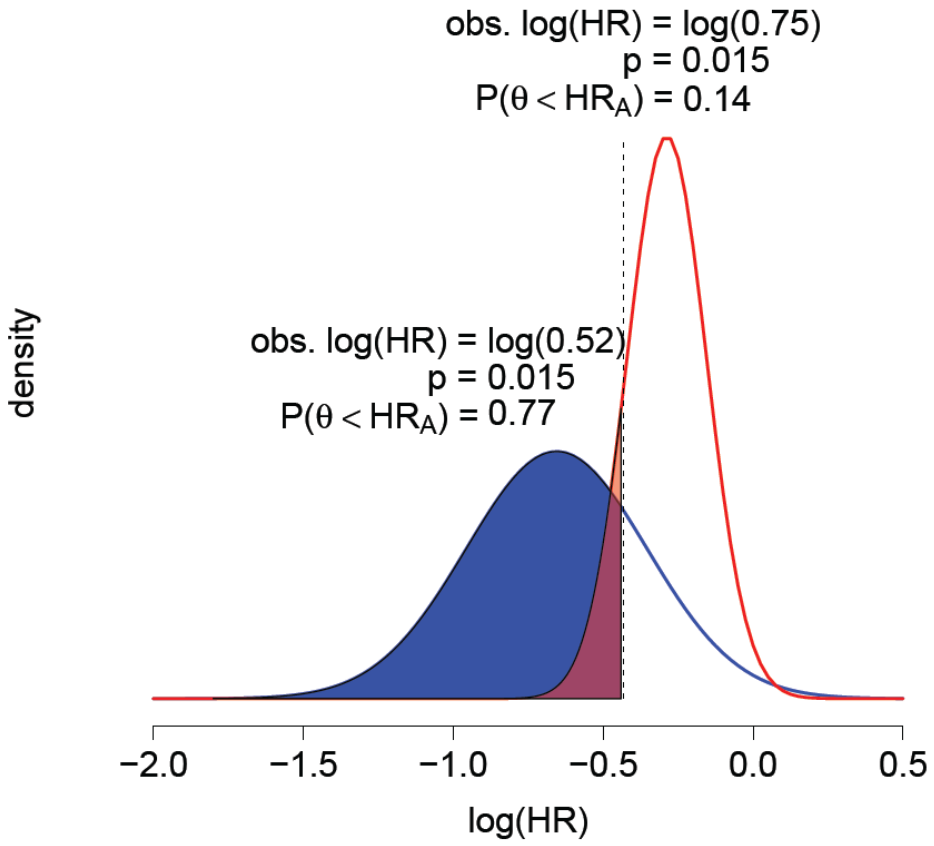
Example (3/4): Bayes factors



- both studies show a significant effect (identical p-values)
- favor HR_A=0.65 or HR_0=1? Bayes factor: 8.02; 5.90

NOVARTIS ONCOLOGY

Example (4/4): (posterior) probabilities

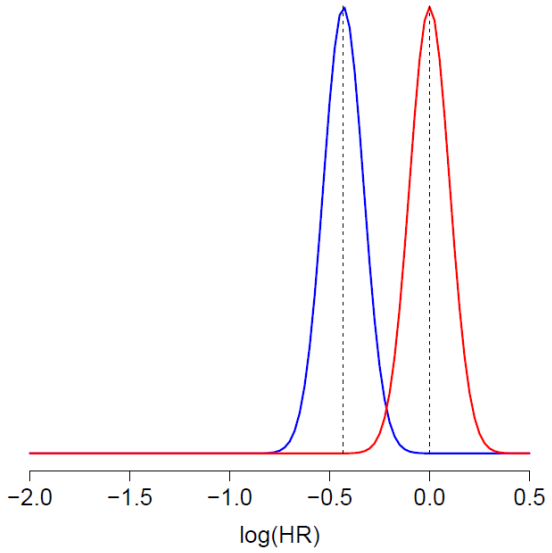


- both studies show a significant effect (identical p-values)
- favor HR_A=0.65 or HR_0=1? Bayes factor: 8.02; 5.90; prob. 0.77; 0.14

NOVARTIS ONCOLOGY

More sophisticated hypotheses

- We compared point hypotheses H_A HR = 0.65 vs H_0 HR = 1
- Yet, we could compare more complex hypotheses
  - can numerically solve intergral for the Bayes factor
  - how to define theses hypotheses?



NOVARTIS ONCOLOGY

A proposal for revised standards (1/3)

- Revised standards need to build on two requirements (Neuenschwander et al 2011)
  - solid evidence for a clinically relevant effect (requires expert input)
  - exclusion of a null-effect
- Requirements can be assessed in classical/Bayesian way
  - classical
    - point estimate above threshold
    - statistical significance
  - Bayesian
    - posterior median above threshold
    - 1-α probability that effect is above null
- This standard can be applied to any (meta-)analysis

A proposal for revised standards (2/3)

- Example: Vilazodone for major depressive disorder
  - two phase III studies
    - primary endpoint: 8-weeks change from baseline in Montgomery-Asberg Depression Rating Scale (MADRS)
    - second study was designed directly on MADRS scale

    The study was planned to enroll 408 patients with 266 patients randomized (133 per arm) to detect a 4.0 difference with a standard deviation of 10

  - 90% power and 2.5% one-sided test provides a critical value of -2.4 (test vs control)
  - we assume **this** (not -4.0) is the minimally clinically relevant difference

A proposal for revised standards (3/3)

- A stringent success criterion
  - observed difference (point estimate) ≤ -2.4
  - statistical significance
- Or equivalently (improper prior)
  - posterior median ≤ -2.4
  - 97.5% posterior probability effect < 0

Study 1

Table 3. Study CLDA-07-DP-02: Sponsor's primary efficacy results: change from baseline to week … ITT sample

		Vilazodone
Sample size		232
Baseline MADRS		
Mean (St		31.9 (3.5)
Median (		32 (22 – 42)
Change fr		
LS Means	-10.8	-13.3
Difference from placebo (SE)		-2.5 (0.96)
(95% confidence interval)		(-4.4, -0.6)
P-value		0.009

-2.5 (se = 0.96)
n = 463
planned: 470

Study 2

Table 10. Study GNSC-04-DP-02: Sponsor's primary analysis: change from baseline to week 8 in … sample

		Vilazodone
Sample size		198
Baseline MADRS		
Mean (Standard		30.8 (3.9)
Median (Min – M		31 (21 – 43)
Change from base		
LS Means	-9.7	-12.9
Difference from placebo (SE)		-3.2 (0.99)
(95% confidence interval)		(-5.1, -1.2)
P-value		0.001

(Source: GNSC-04-DP-02 Study Report; Tables 11-6 & 11-7, page 66)

-3.2 (se = 0.99)
n = 397
planned: 266

NOVARTIS ONCOLOGY

Agenda

Background

P-values, Bayes factors and probabilities

A holistic view of evidence

Bayesian evidence synthesis: case study

Conclusion

Evidence assessment

On the one hand, the CA209-066 study included only patients with BRAF V600 wt tumour, ***which, accordingly, did not concur with patients of the research question.***
([A15-27] Nivolumab – Benefit assessment according to § 35a Social Code Book V (dossier assessment)

Sufficient relevant evidence?

The manufacturer's decision problem was ***substantially narrower than that of the NICE scope****, primarily in terms of the population considered.*
(Vortioxetine for treating major depressive disorder)

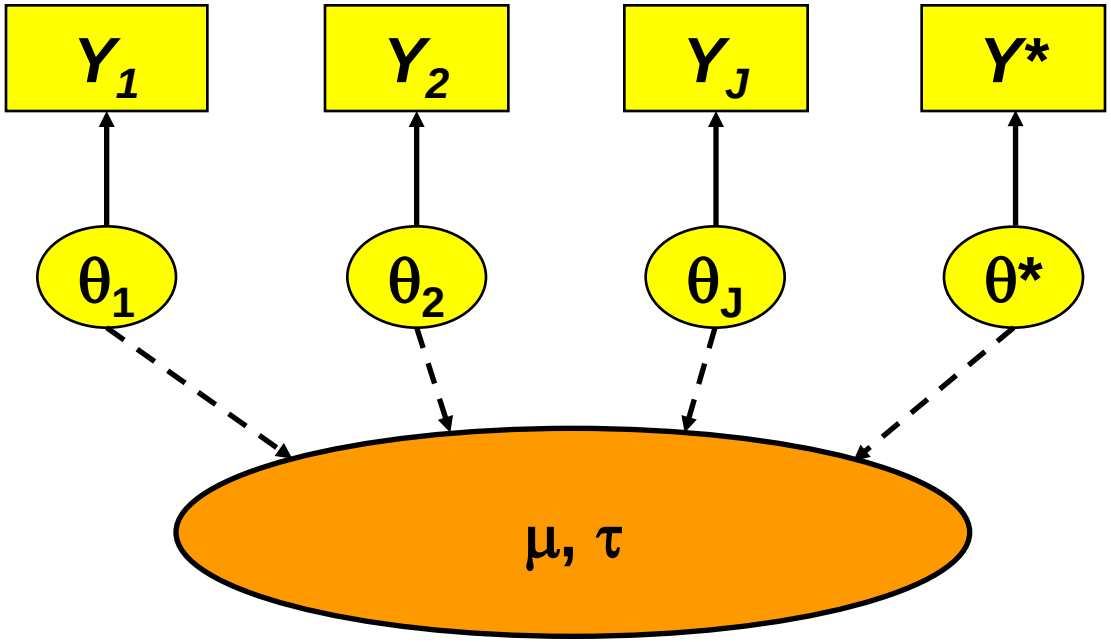
The applicant ... demonstrated the efficacy of vemurafenib primarily based on Study NO25026 ... ***Study NP22657 as supportive evidence****. [...] The efficacy of vemurafenib demonstrated in Study NO25026* ***was supported by the findings in Study NP22657***
CDER statistical review for zelboraf (vemurafenib)

Benefit assessment

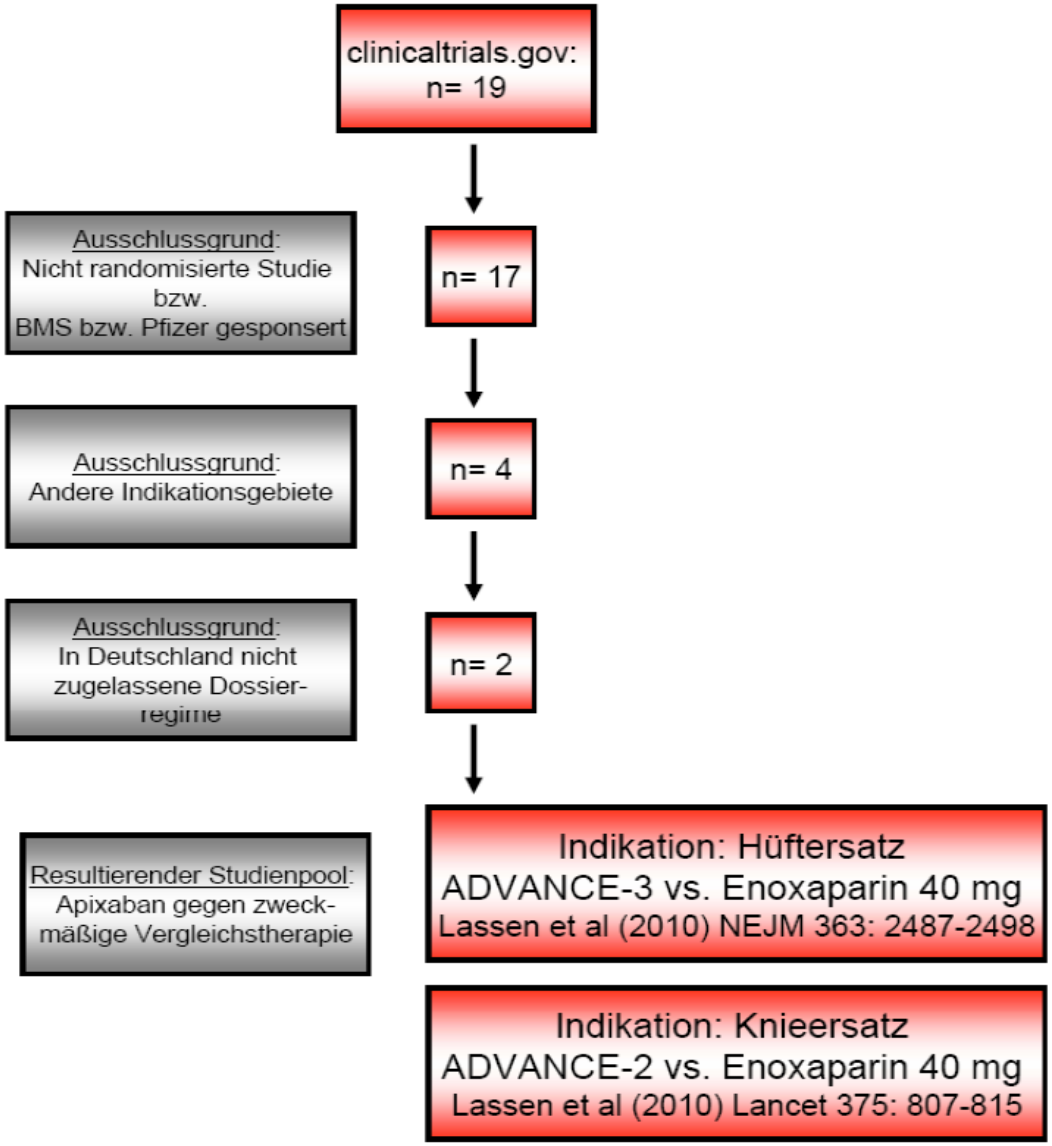
- Benefit-assessment is inherently challenging
  - typical clinical (registration) study population is **not** a random sample of the target population
    - specialized centers / investigators
    - participation is an individual choice influenced by patient-related factors (e.g. Longtin et al, 2010)
  - treatment effect estimates could be
    - too optimistic: better adherence to medication, better oversight, etc
    - too pessimistic: better outcomes in control group (Penny et al, 2016)
    - not directly interpretable: different standard of care
- Current trends
  - use of real-world-evidence data to assess population benefit-risk
  - modeling natural disease history to assess population impact

Uncertainty assessment

- Normal-normal hierarchical model
  - often we are interested in the population mean μ
  - yet this does not fully reflect all uncertainty
  - role of the prediction interval (e.g. Guddat et al 2012)



Apixaban to prevent vn thromboembolism (1/6)

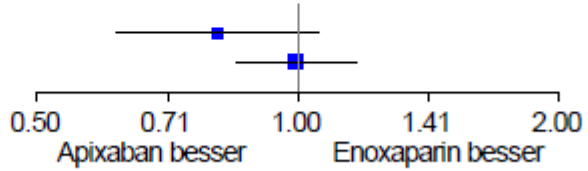


Apixaban to prevent vn thromboembolism (2/6)

- Very narrow definition, yet
  - heterogeneity may still be present
  - answer = no meta-analysis?

Unerwünschte Ereignisse: Blutungen, Behandlungsperiode
Modell mit zufälligen Effekten - DerSimonian und Laird (zur Darstellung der Gewichte)

Studie	Apixaban n/N	Enoxaparin n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
ADVANCE-2	90/1501	112/1508		36.2	0.81	[0.62, 1.06]
ADVANCE-3	268/2673	268/2659		63.8	0.99	[0.85, 1.17]



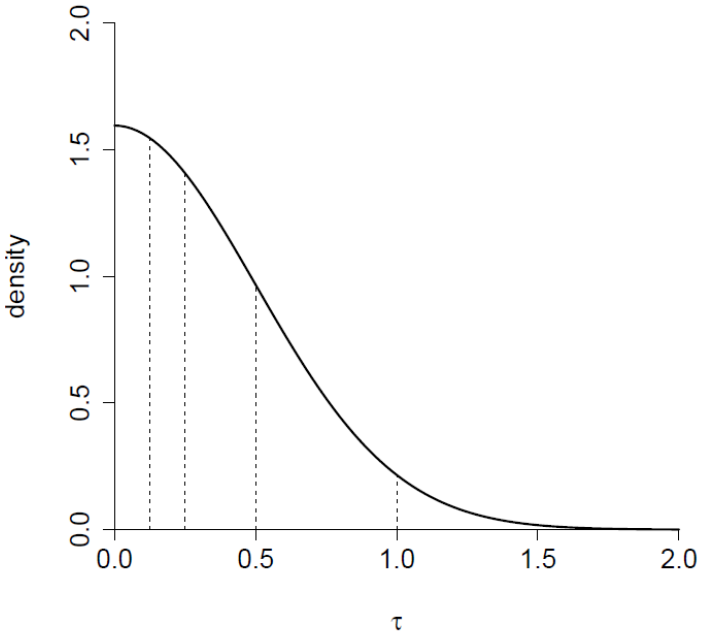
Heterogenität: Q=1.72, df=1, p=0.190, I^2=41.7%

  - prediction interval reflects uncertainty
    - difficult with few studies in classical approach (e.g. Friede et al, to appear)
    - Bayesian: straightforward; HN(0.5) prior for τ (e.g. Friede et al, to appear)

NOVARTIS ONCOLOGY

Apixaban to prevent vn thromboembolism (3/6)

- Interpretation of τ
  - related to outcome scale
  - for log-risk-ratio, HN(0.5)
    - range: small to large heterogeneity (95% interval: 0.02; 1.12)
    - median (0.34) moderate-to-substantial heterogeneity
    - ratio of risk ratios (97.5% to 50%): ~ 3.00



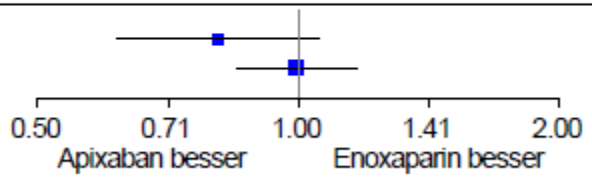
Heterogeneity	σ/τ	τ (σ=2)	exp ($\theta_{97.5\%}$)/exp($\theta_{50\%}$)
large	2	1	7.10
substantial	4	0.5	2.66
moderate	8	0.25	1.63
small	16	0.125	1.28

NOVARTIS ONCOLOGY

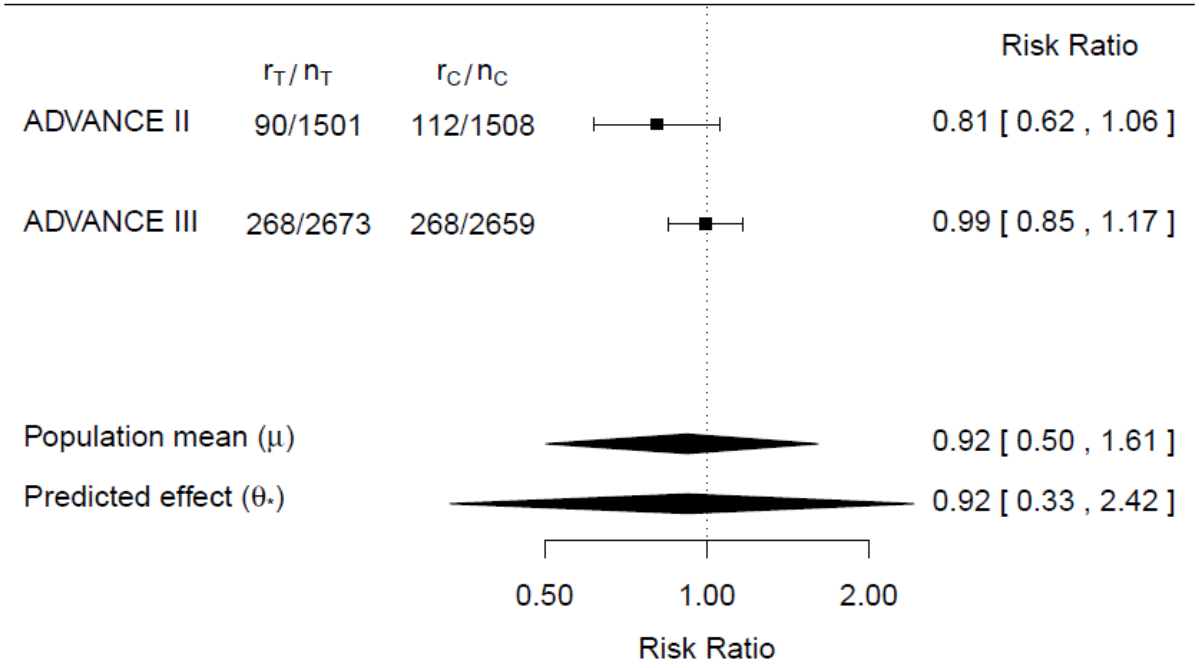
Apixaban to prevent vn thromboembolism (4/6)

Unerwünschte Ereignisse: Blutungen, Behandlungsperiode
Modell mit zufälligen Effekten - DerSimonian und Laird (zur Darstellung der Gewichte)

Studie	Apixaban n/N	Enoxaparin n/N	RR (95%-KI)	Gewichtung	RR	95%-KI
ADVANCE-2	90/1501	112/1508		36.2	0.81	[0.62, 1.06]
ADVANCE-3	268/2673	268/2659		63.8	0.99	[0.85, 1.17]



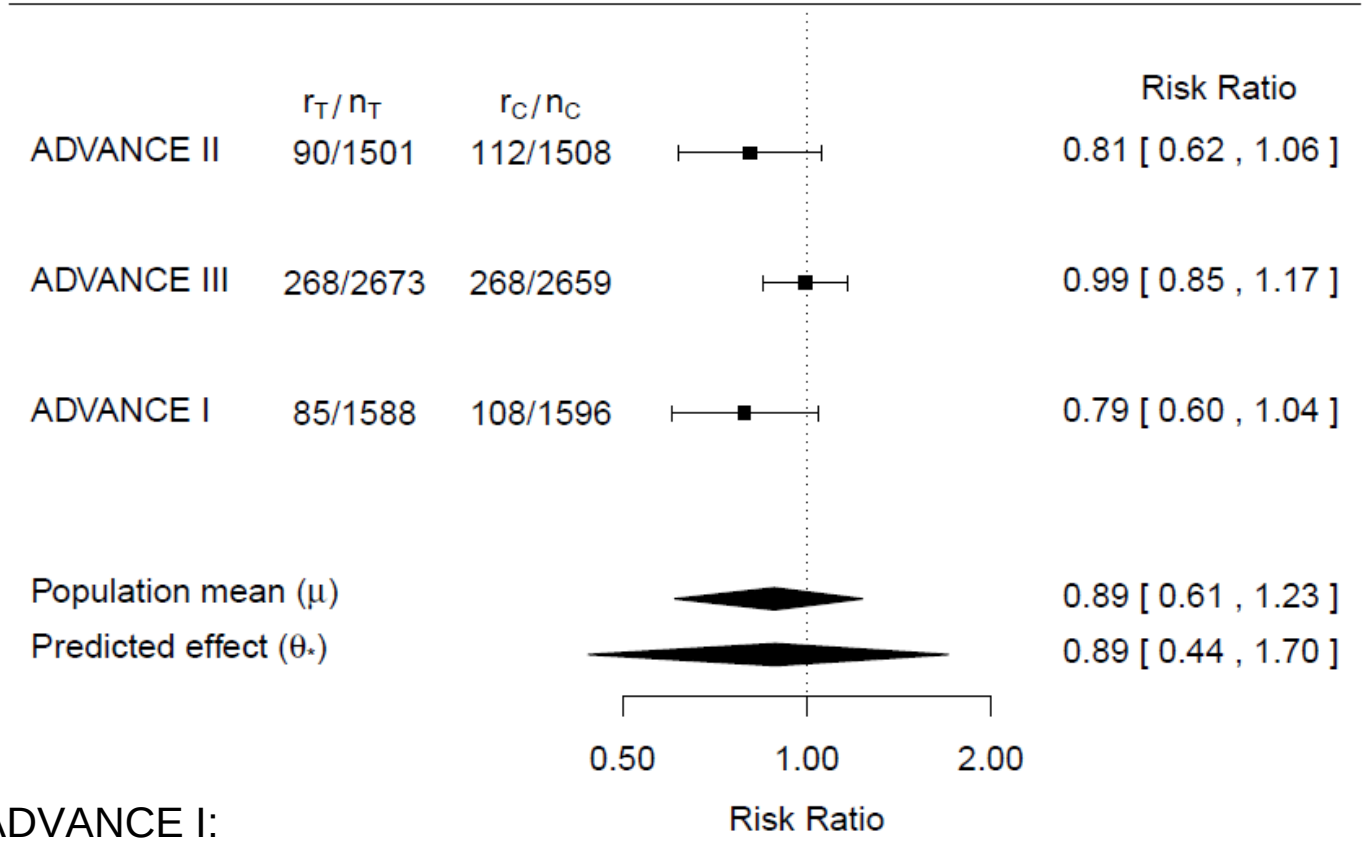
Heterogenität: Q=1.72, df=1, p=0.190, I^2=41.7%



NOVARTIS ONCOLOGY

Apixaban to prevent vn thromboembolism (5/6)

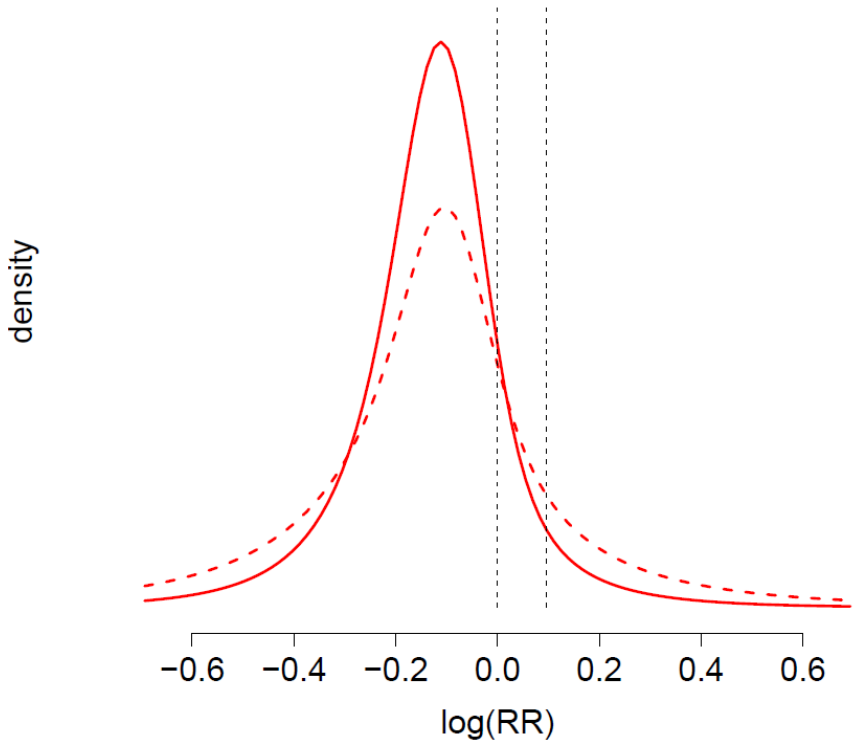
- Why not include additional studies (sensitivity)?



- ADVANCE I: https://clinicaltrials.gov/ct2/show/results/NCT00371683?sect=X01256#all

Apixaban to prevent vn thromboembolism (6/6)

- If we are interested in risk quantification – double criterion
  - probability to be worse (risk ratio > 1)
  - probability to be above some relevant threshold (**arbitrarily** RR = 1.1)
  - population mean (3 studies)
    - P(RR > 1) = 0.15
    - P(RR > 1.1) = 0.06
  - predicted effect (3 studies)
    - P(RR* > 1) = 0.24
    - P(RR* > 1.1) = 0.14



NOVARTIS ONCOLOGY

Agenda

Background

P-values, Bayes factors and probabilities

A holistic view of evidence

Case study: Bayesian evidence synthesis

Conclusion

Example: Zirgan for herpetic keratitis (1/4)

- Herpetic keratitis (Kaye et al; White et al; Dawson et al; Suresh et al)
  - inflammatory condition of the eye caused by the herpes simplex virus
  - leading cause of corneal blindness in the industrialized world
  - orphan disease
- 2008: manufacturer seeks FDA approval for Zirgan 0.15%
  - not the first treatment – acyclovir effective, yet potential side effects
  - three (!) phase II and one phase III study
  - goal: to establish non-inferiority
  - our focus: cure rate at day 14
  - for more information: see FDA's approval documents and Wandel et al (in preparation)

Example: Zirgan for herpetic keratitis (2/4)

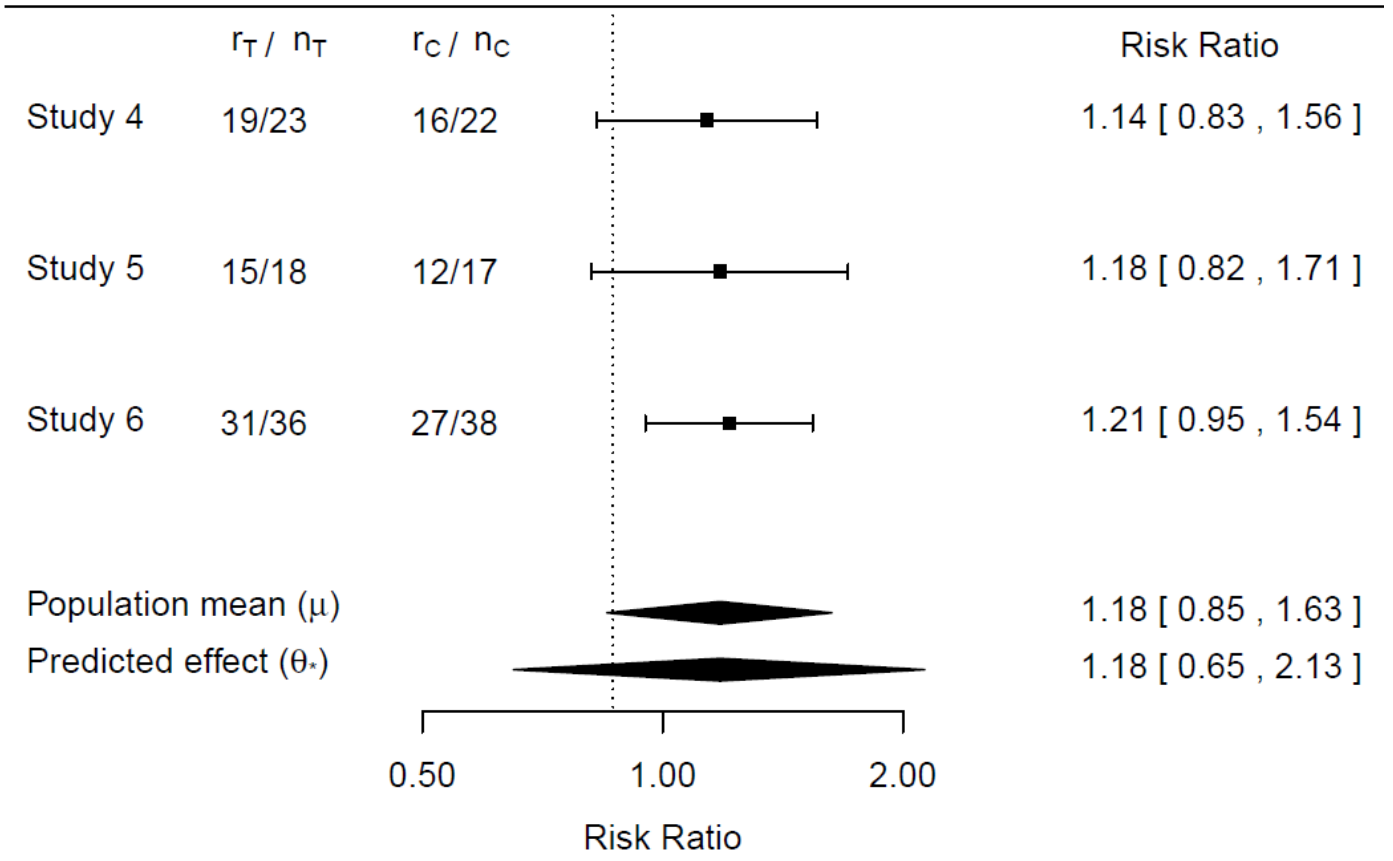
Table 2. Data of Phase II and III studies

	Study (Phase)			
	4 (II)	5 (II)	6 (II)	7 (III)
Objective	Efficacy & Safety	Efficacy & Safety	Efficacy & Safety	Efficacy & Safety
Design	3-arm randomized	2-arm randomized	3-arm randomized	2-arm randomized
Location	Africa	Europe	Pakistan	Europe & Africa
Product	G: 0.15%, 0.05%; A: 3%	G: 0.15%; A: 3%	G: 0.15%, 0.05%; A: 3%	G: 0.15%; A: 3%
Regimen	1	1	2	1
Study period (months)	4/90–5/92 (25)	12/90–5/92 (18)	5/91–10/92 (18)	9/92–9/94 (25)
Total cure rate, day 14 (%)				
G 0.15%	19/23 (82.6)	15/18 (83.3)	31/36 (86.1)	
A 3%	16/22 (72.7)	12/17 (70.6)	27/38 (71.1)	

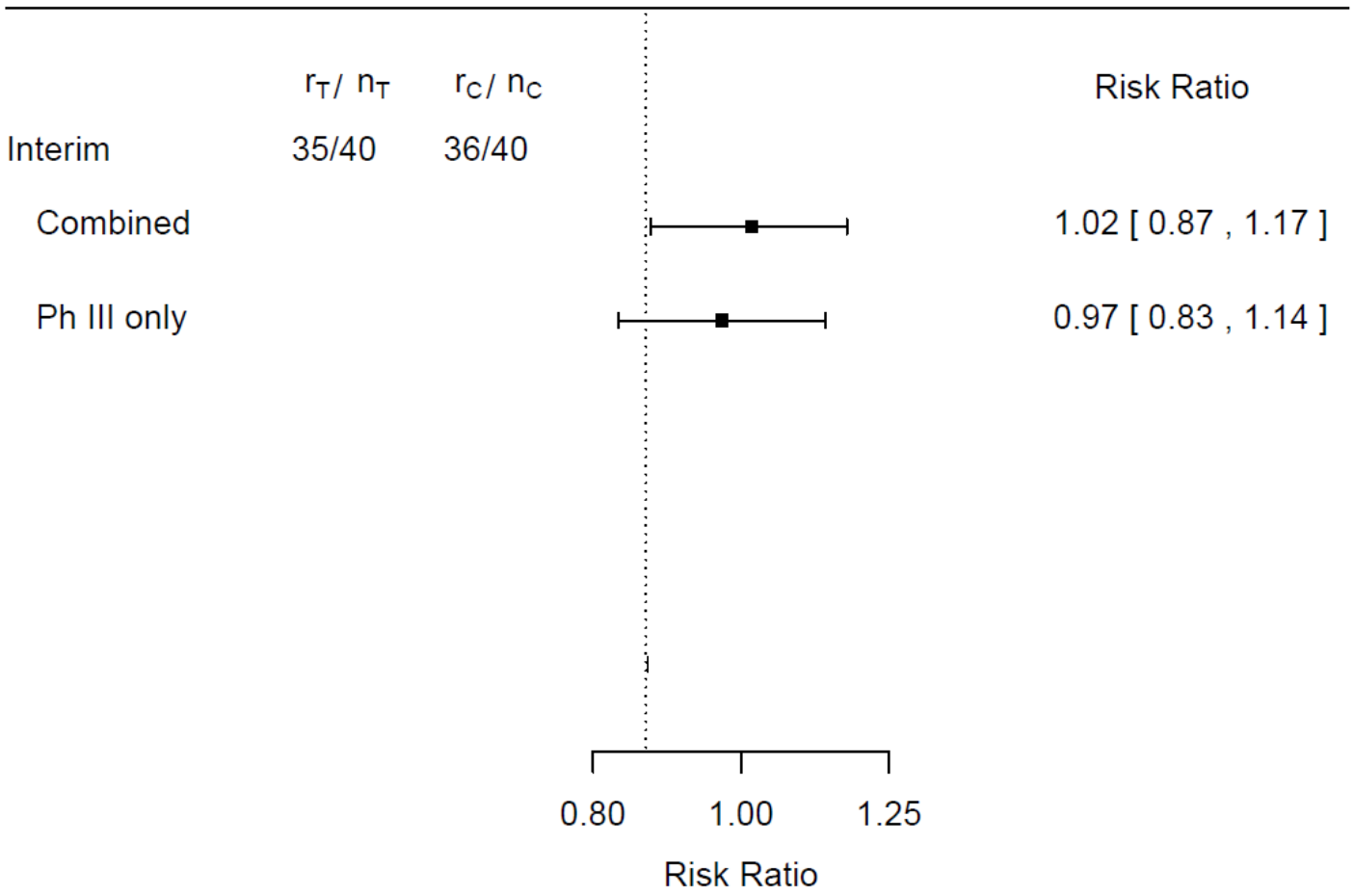
G = Ganciclovir (Ganciclovir 0.15% = Zirgan), A = Acyclovir
Regimen: 1 = 1 drop 5x/day until ulcer healed, then 1 drop 3x/day for 7 days; 2 = 1 drop 5x/day for 10 days

- Potential options?
  - recognize ph II evidence but ignore for statistical evaluation (FDA)
  - perform a meta-analysis of all data (classical/Bayesian)
  - phase II data as prior information for the phase III analysis (Bayesian)

Example: Zirgan for herpetic keratitis (3/4)



Example: Zirgan for herpetic keratitis (4/4)



NOVARTIS ONCOLOGY

Agenda

Background

P-values, Bayes factors and probabilities

A holistic view of evidence

Case study: Bayesian evidence synthesis

Conclusion

The world is changing

VICE PRESIDENT JOE BIDEN'S MOONSHOT ADDRESS

[...] break down barriers to progress by promoting data sharing and facilitating collaborations [...]

Vice President Joe Biden **gave remarks June 6** on the White House's Cancer Moonshot Initiative to accelerate cancer research efforts and break down barriers to progress by promoting data sharing and facilitating collaborations to advance cancer prevention, treatment, and care. The Moonshot Initiative, like ASCO's Annual Meeting and Multi-disciplinary Thematic Symposia, is committed to bringing the collective cancer community together to share knowledge to improve patient care.

http://am.asco.org/virtual-meeting-on-demand

- source: http://am.asco.org/virtual-meeting-on-demand/presentation/biden-info

NOVARTIS ONCOLOGY

Improved standards for statistical evidence (1/2)

- A holistic view of evidence is needed
  - we may not always have «perfect» data at hand
  - yet imperfect data may be very informative and supportive
  - as statisticians, it is our task to deal with uncertainty
- Approaches to borrowing depend on heterogeneity
  - Meta-analytic (Schmidli et al 2014; Neuenschwander et al 2016)
  - Robust versions (Schmidli et al 2014; Leon-Novelo 2013)
- More data – when used correctly - will lead to better decisions
  - yet uncertainty may decrease or increase
- Decision making goes beyond statistics!

Improved standards for statistical evidence (2/2)

The fact that network meta-analyses have become so popular is not surprising ***because they answer the real questions of interest to decision makers****, who are usually faced with an array of treatment options, not just two.*

Higgins J, Welton N. Network meta-analysis: a norm for comparative eff ectiveness? Lancet (2016)

Thank you for your attention

NOVARTIS ONCOLOGY

Acknowledgments

- Satrajit Roychoudhury
- Beat Neuenschwander
- Tim Friede
- Christian Röver

References: publications (1/2)

- Baker M. *Is there a reproducibility crisis?* Nature 2016; 533:452-4.
- Dawson C and Togni B. *Herpes simplex eye infections: clinical manifestations, pathogenesis and management.* Surv Ophthalmol 1976; 21:121-35.
- Friede T, Röver C, Wandel S, Neuenschwander B. *Meta-analysis of few small studies in orphan diseases.* To appear in: Res Synth Methods
- Friede T, Röver C, Wandel S, Neuenschwander B. *Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases* accepted: Biometrical Journal.
- Guddat C, Grouven U, Bender R, Skipka G. *A note on the graphical presentation of prediction intervals in random-effects meta-analyses.* Syst Rev. 2012;1:34.
- Higgins J, Welton N. *Network meta-analysis: a norm for comparative effectiveness?* Lancet. 2015;386:628-30.
- Johnson VE. *Revised standards for statistical evidence.* PNAS 2013; 26;110:19313-7
- Johnson VE. *Uniformly most powerful Bayesian tests.* Ann Stat. 2013; 41:1716-41
- Kaye S and Choudhary A. *Herpes simplex keratitis.* Prog Retin Eye Res 2006; 25:355-80.
- Leon-Novelo L, Bekele B, Müller P et al. *Borrowing strength with nonexchangeable priors over subpopulations.* Biometrics. 2012;68:550-8.

References: publications (2/2)

- Longtin Y, Sax H, Leape L et al. *Patient participation: current knowledge and applicability to patient safety.* Mayo Clin Proc. 2010;85:53-62.
- Neuenschwander B, Rouyrre N, Hollaender N et al. *A proof of concept phase II non-inferiority criterion.* Stat Med. 2011;30:1618-27.
- Neuenschwander B, Roychoudhury S, Schmidli H. *On the use of co-data in clinical trials*. Statistics in Biopharmaceutical Research (2016)
- Schmidli H, Gsteiger S, Roychoudhury S et al. *Robust meta-analytic-predictive priors in clinical trials with historical control information.* Biometrics 2014; 70:1023-32.
- Penny M, Verity R, Bever C et al. *Public health impact and cost-effectiveness of the RTS,S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models.* Lancet. 2016;387:367-75.
- Suresh P and Tullo A. *Herpes simplex keratitis.* Indian J Ophthalmol 1999; 47:155-165.
- Wandel S, Neuenschwander B, Röver C, Friede T. *Using phase II data for the analysis of phase III studies: an application in rare diseases.* (in preparation)
- Wasserstein R, Lazar N. *The ASA's statement on p-values: context, process, and purpose.* The American Statistician (2016)
- White ML and Chodosh J. Herpes Simplex Virus Keratitis: A Treatment Guideline. In: Hoskins Centers Compendium of Evidence-Based Eye Care

References: review documents

- Apixaban benefit assessment: https://www.g-ba.de/downloads/92-975-72/2011-06-15-D-009_Apixaban_IQWiG-Nutzenbewertung.pdf
- Nivolumab benefit assessment (IQWiG): https://www.iqwig.de/en/projects-results/projects/drug-assessment/a15-27-nivolumab-benefit-assessment-according-to-35a-social-code-book-v-dossier-assessment.6891.html
- Vilazodone statistical review (FDA): http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022567Orig1s000StatR.pdf
- Vortioxetine cost-benefit assessment (NICE): https://www.nice.org.uk/guidance/TA367/documents/major-depressive-disorder-vortioxetine-id583-committee-papers2
- Zelboraf statistical review (FDA): http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/202429Orig1s000StatR.pdf
- Zirgan approval documents (FDA): http://www.accessdata.fda.gov/drugsatfda_docs/nda/2009/022211_zirgan_toc.cfm

References: websites, R packages

- PubPeer: https://pubpeer.com/
- Retraction Watch: http://retractionwatch.com/
- bayesmeta: https://cran.r-project.org/web/packages/bayesmeta/index.html
- metafor: https://cran.r-project.org/web/packages/metafor/index.html