No comparators no problem?

Case study of Entrectinib

Iain Bennett
Evidence Enabler, Global Access

Disclaimer/conflict of interest

- I am an employee and shareholder of Hoffmann La Roche AG.
- The views and opinions presented here are my own and do not necessarily reflect those of Roche.

1982 2019

1982

NTRK mutation first identified

2019

EMA approval for first TRK inhibitor

Reference: A Vaishnavi, AT Le, RC Doebele; TRKing down an old oncogene in a new era of targeted therapy; Cancer Discov, 5 (2015), pp. 25-34. https://doi.org/10.1158/2159-8290.CD-14-0765
https://www.ema.europa.eu/en/news/first-histology-independent-treatment-solid-tumours-specific-gene-mutation#:~:text=EMA's%20human%20medicines%20committee%20(CHMP,Kinase%20(NTRK)%20gene%20fusion.

48.3 years

48.3 years

Minimum time to results in years for an RCT in Tumor Agnostic NTRK

Reference: Lozeno-Ortega et al; “TUMOUR-SPECIFIC RANDOMIZED CONTROLLED TRIALS IN RARE ONCOGENE-DRIVEN CANCERS- ASKING FOR THE IMPOSSIBLE?”; ISPOR Europe 2019; https://www.ispor.org/heor-resources/presentations-database/presentation/euro2019-3122/96780

17 years

17 years

Minimum time to results in years for an RCT in any NTRK sub indication (Sarcoma)

Reference: Lozeno-Ortega et al; “TUMOUR-SPECIFIC RANDOMIZED CONTROLLED TRIALS IN RARE ONCOGENE-DRIVEN CANCERS- ASKING FOR THE IMPOSSIBLE?”; ISPOR Europe 2019; https://www.ispor.org/heor-resources/presentations-database/presentation/euro2019-3122/96780

36%

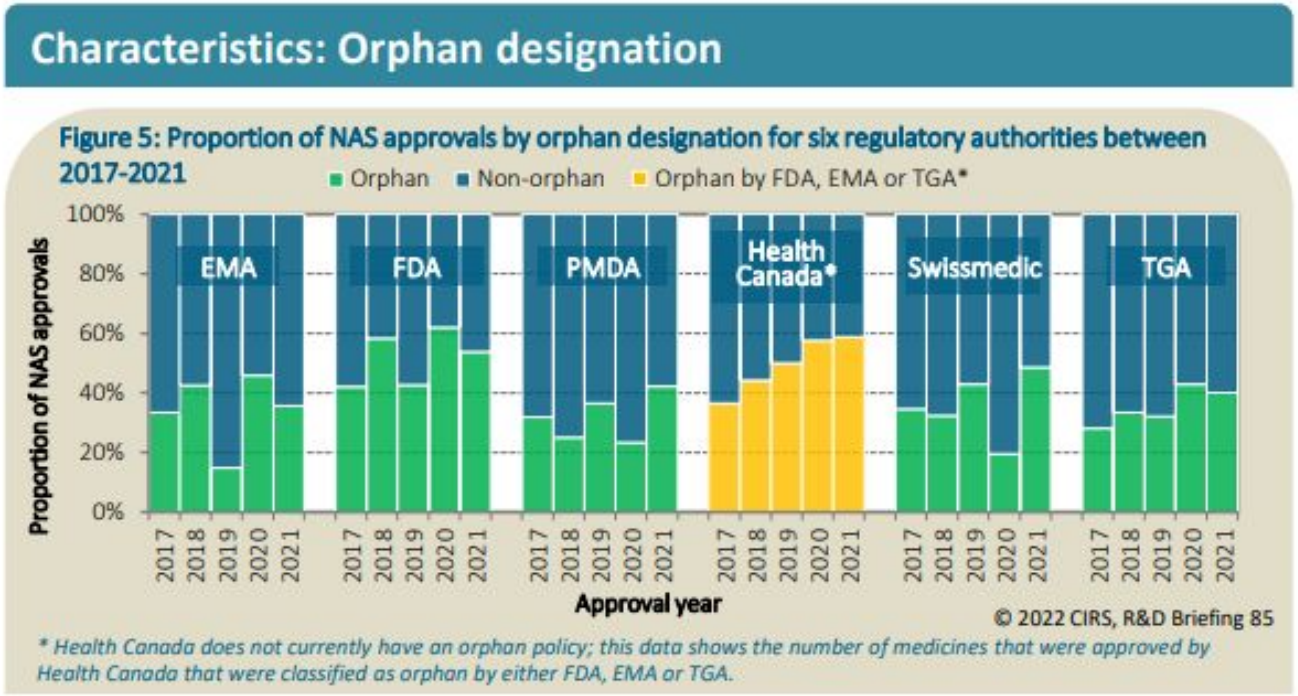
36%

Percent of EMA approvals in 2021 that had orphan designation

Reference: Centre for Innovation in Regulatory Science (2022) R&D Briefing 85: New drug approvals in six major authorities 2012–2021: Focus on Facilitated Regulatory Pathways and internationalisation. Centre for Innovation in Regulatory Science (CIRS), London, UK.

Context: how common is rare?

59% of Approvals by Health Canada in 2021 had orphan designation by EMA, FDA, or TGA



Reference: Centre for Innovation in Regulatory Science (2022) R&D Briefing 85: New drug approvals in six major authorities 2012–2021: Focus on Facilitated Regulatory Pathways and internationalisation. Centre for Innovation in Regulatory Science (CIRS), London, UK.

Contents

Case study background

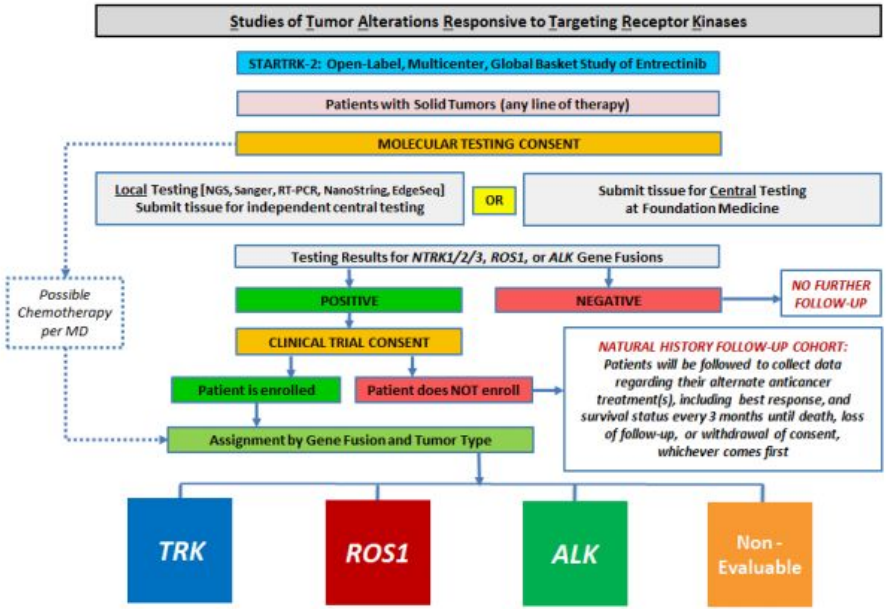
Trial design

STARTRK-2

STARTRK-2 trial characteristics

- Phase II, open label
- **Single arm**
- Entrectinib as intervention
- Basket trial with
  - NTRK basket Tumor site agnostic
  - ROS1 basket mNSCLC
- Tumors assessed by blinded independent central review (BICR) using RECIST version 1.1

Figure 2 *GO40782* **(STARTRK-2) Basket Study Schema**



Reference: Doebele RC, Drilon A, Paz-Ares L, et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. Lancet Oncol. 2020;21(2):271-282

Drilon A, Siena S, Dziadziuszko R, et al. Entrectinib in ROS1 fusion-positive non-small-cell lung cancer: integrated analysis of three phase 1–2 trials. Lancet Oncol. 2020;21(2): 261–270

Two populations

NTRK

NTRK mutation

- Tumor site agnostic
- 0.3% of all solid tumors
- No orphan status

ROS1

ROS1 mutation

- mNSCLC
- 1-2% of metastatic lung cancers
- No orphan status

Reference: Doebele RC, Drilon A, Paz-Ares L, et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. Lancet Oncol. 2020;21(2):271-282
Drilon A, Siena S, Dziadziuszko R, et al. Entrectinib in ROS1 fusion-positive non-small-cell lung cancer: integrated analysis of three phase 1–2 trials. Lancet Oncol. 2020;21(2): 261–270

Regulators and HTA - different hats different questions

HTA needs to answer two important questions

Among many others…

Comparative effectiveness

What are the effects (benefits/risks) of **this intervention compared to other** available options?

Generalisability/external validity

What are the effects of this intervention in **my population**?

HTA needs to answer two important questions

Among many others…

Comparative effectiveness

What are the effects (benefits/risks) of **this intervention compared to other** available options?

Decision problem: Should a city buy beach umbrellas?

efficacy:
beach umbrellas block sun

HTA needs to answer two important questions

Among many others…

Comparative effectiveness

What are the effects (benefits/risks) of **this intervention compared to other** available options?

Decision problem: Should a city buy beach umbrellas?

efficacy:
beach umbrellas block sun

comparative effectiveness:
beach umbrella vs t-shirt:

- How much more protection?

HTA needs to answer two important questions

Among many others…

Comparative effectiveness

What are the effects (benefits/risks) of **this intervention compared to other** available options?

Generalisability/external validity

What are the effects of this intervention in **my population**?

Decision problem: Should a city buy beach umbrellas?

efficacy:
beach umbrellas block sun

comparative effectiveness:
beach umbrella vs t-shirt:
- How much more protection?

context:

Sydney, Australia
- 2628 hours of sun/year

x% more protection → *y* hours protection

Basel, Switzerland
- 1640 hours of sun/year

→ *z* hours protection

Question 1: Comparative effectiveness

Comparative effectiveness

How can we do this with a single arm trial?

Some options:

1) A priori thresholds
2) Intra-patient comparisons
3) External control (other clinical trials)
4) External control (real world data)
5) Real world analysis

NTRK
ROS1

1) A priori thresholds

In STARTRK-2 a null hypothesis and sample size were defined for Primary Endpoint (ORR)

NTRK

Null hypothesis: ORR **<= 20%**

Power 80%, Alpha 5% (one sided 2.5%)

"The choice of a statistically significant observed response rate of > 20% for locally advanced or metastatic solid tumor gene rearrangement baskets is based upon a review of the literature of expected response rates to standard, non-targeted therapies in these diseases."

ROS1

Null hypothesis: ORR **<= 50%**

Power 80%, Alpha 5% (one sided 2.5%)

"The choice of a statistically significant observed response rate of >50% for this population is based on review of the literature of expected response rate to available targeted therapy, i.e., crizotinib [Shaw et al, 2014]."

Reference: STARTRK-2 Protocol (Lancet Oncology supplementary appendix)
Note: This is a simplified view as both baskets actually used a sequential testing procedure and staggered enrollment while maintaining these design characteristics.

2) Intra-patient comparisons

Prior line of therapy results as a proxy for Standard of Care

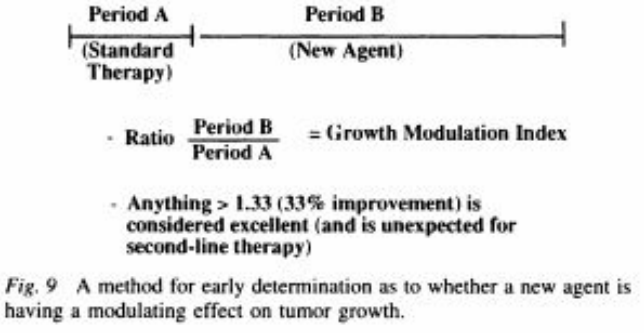
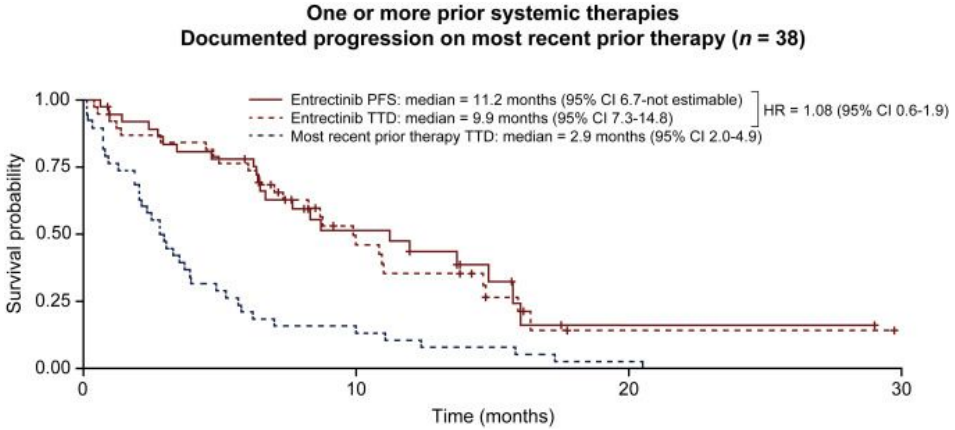


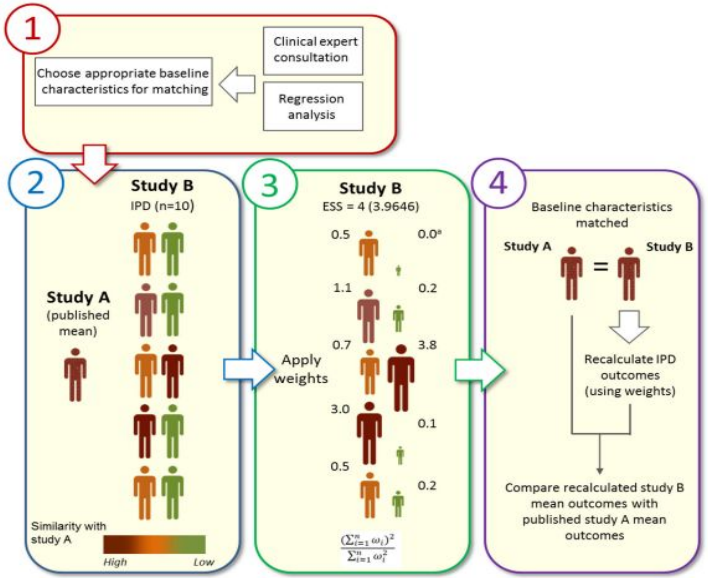
Fig. 9 A method for early determination as to whether a new agent is having a modulating effect on tumor growth.



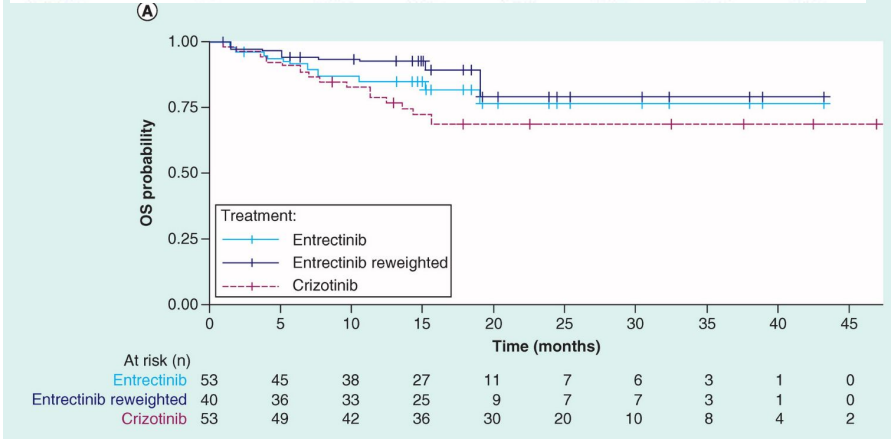
Reference: Krebs M et al. Intrapatient comparisons of efficacy in a single-arm trial of entrectinib in tumour-agnostic indications. ESMO Open. 2021 Apr;6(2):100072. doi: 10.1016/j.esmoop.2021.100072.
Illustration: Von Hoff DD. There are no bad anticancer agents, only bad clinical trial designs–twenty-first Richard and Hinda Rosenthal Foundation Award Lecture. Clin Cancer Res. 1998;4(5):1079-1086.

3) External control (other clinical trials)

Reweight patients to match target (**other** trial) population characteristics



Intervention	Sample size (effective sample size)	Female (%)	ECOG 2 (%)	Never smoked (%)	Age (yr)	Treatment naive (%)	Stage IV – CNS (%)
Entrectinib	53	64.15	11.32	58.49	53.55	13.21	43.40
Entrectinib reweighted	53 (34.18)	56.60	1.89	75.47	55.00	13.21	18.10
Crizotinib	53	56.60	1.89	75.47	55.00	13.21	18.10



Reference: Chu P et al. Matching-adjusted indirect comparison: entrectinib versus crizotinib in ROS1 fusion-positive non-small cell lung cancer. Journal of Comparative Effectiveness Research 2020; 9(15): 861-876 https://doi.org/10.2217/cer-2020-0063

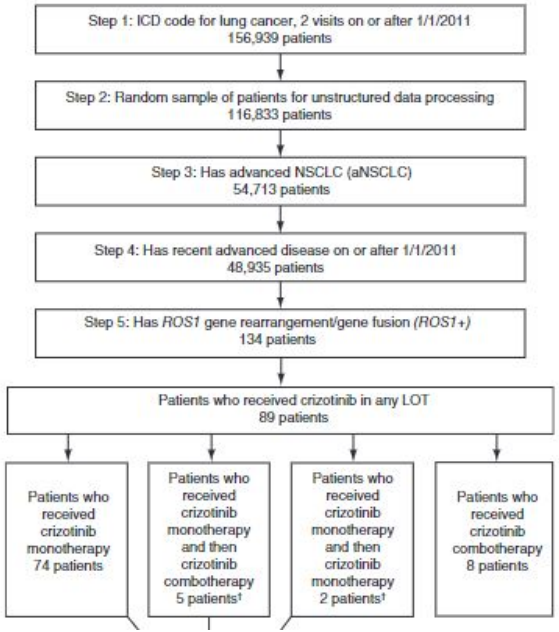
R code: https://github.com/Roche/maic

Illustration: Nash P, et al, Secukinumab Versus Adalimumab for Psoriatic Arthritis: Comparative Effectiveness up to 48 Weeks Using a Matching-Adjusted Indirect Comparison. Rheumatol Ther (2018) 5:99–122

4) External control (real world data)

Filter & reweight patients to match target (**your** trial) population characteristics

156,939



65

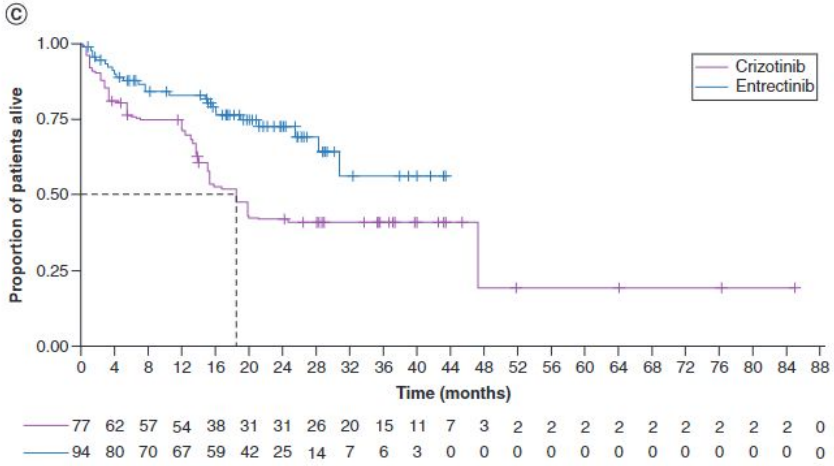


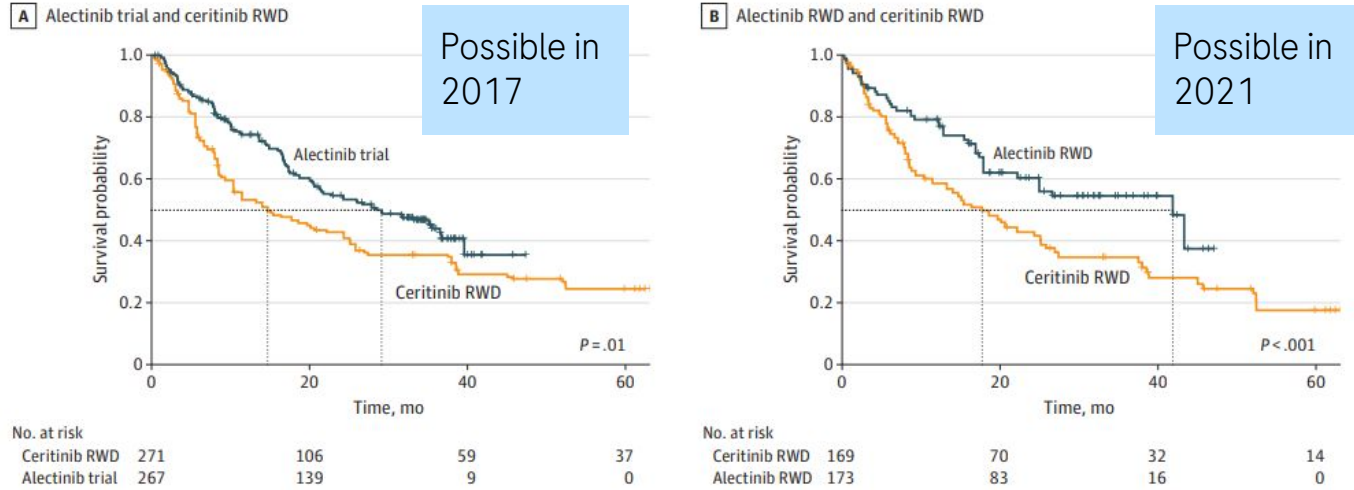
Figure 2. Kaplan–Meier estimates of the primary and secondary endpoints in the crizotinib and entrectinib cohorts.
(A) TTD, **(B)** PFS and **(C)** OS. All curves are weighted based on the propensity score for comparability between arms.
HR: Hazard ratio; OS: Overall survival; PFS: Progression-free survival; TTD: Time-to-treatment discontinuation.

Reference: Doebele RC et al. Comparative effectiveness analysis between entrectinib clinical trial and crizotinib real-world data in ROS1+ NSCLC. J. Comp. Eff. Res.(2021) 10(17). Doi: .

5) Real world analysis

Wait for more data

Figure 1. Adjusted Kaplan-Meier Curves for Overall Survival



Curves show survival for alectinib trial data vs ceritinib real-world data (RWD) (A) and alectinib RWD vs ceritinib RWD (B). Numbers at risk are reweighted sample sizes. Log-rank *P* values are shown.

Reference: Wilkinson et al; Assessment of Alectinib vs Ceritinib in ALK-Positive NSCLC in Phase 2 Trials and Real-world Data; JAMA Network Open. 2021;4(10):e2126306. doi:10.1001/jamanetworkopen.2021.26306

Reference: Davies J, Martinec M, Delmar P, et al. Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib.J Comp Eff Res. 2018;7(9):855-865. doi:10.2217/cer-2018-0032

Question 2: External validity

External validity

How can we do this?

Some options:

1) Clinical extrapolation
2) Subgroups
3) Re-weighting
4) Bayesian hierarchical models

3) Re-weighting

Propensity weighting to generate effectiveness estimates

Table 1. Characteristics of 1,156 HIV-infected Patients in the AIDS Clinical Trial Group 320 Study in 1996–1997 Followed for 1 Year and of the Estimated 54,220 HIV-infected Individuals in the United States in 2006

Characteristic[a]	Trial Patients		US Population	
	No.	%	No.	%
Age, years	38 (33, 44)		NA	
Age group, years[b]				
13–29	106	09	18,500	34
30–39	515	45	16,740	31
40–49	388	34	13,370	25
≥50	147	13	5,610	10
Male sex	956	83	39,810	73
Race				
White, non-Hispanic	623	54	19,580	36
Black, non-Hispanic	328	28	24,920	46
Hispanic	205	18	9,720	18
CD4 count (cells/mm^3)[c]	75 (33, 137)		NA	

ITT

US population weighted

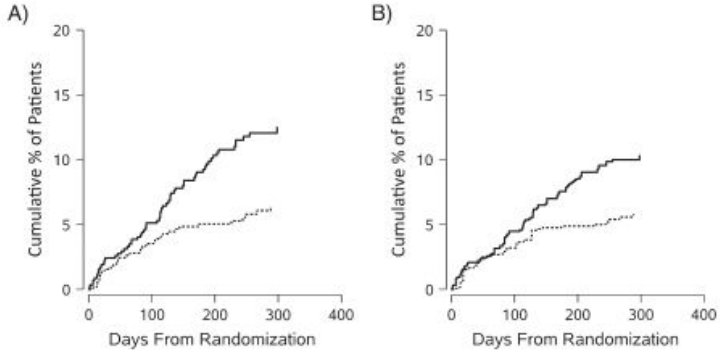


Figure 1. Complement of the Kaplan-Meier survival curves, acquired immunodeficiency syndrome (AIDS) Clinical Trial Group 320 Study, 1996–1997, United States. A) intent-to-treat; B) selection probability weighted. Solid lines represent patients randomly assigned to the control group; dashed lines represent patients randomly assigned to the treatment group.

Reference: Cole SR, Stuart EA; Generalizing Evidence From Randomized Clinical Trials to Target Populations; Am J Epidemiol 2010;172:107–115

Reference: GetReal - Project No. 115546 WP1: Deliverable 1.5/1.6 Case Study: Propensity Weighting and Extrapolation in Non Small Cell Lung Cancer https://rwe-navigator.eu/use-real-world-evidence/model-effectiveness-in-the-real-world/overview-of-methods-for-predicting-outcomes/propensity-weighting-to-generate-estimates-of-relative-effectiveness-from-trials/

4) Bayesian hierarchical models

How to estimate a subgroup you haven't observed?

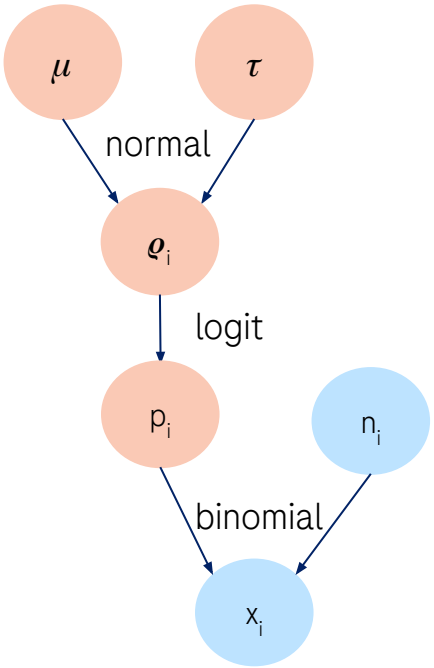


Table 2 Probabilities of response for all histologies

Histology	Observed Response	Estimated Mean Response Based on BHM (%)	95% CrI
Fixed effects			
Pooled	41/55 = 74.5%	74.20%	62.0%–84.7%
Random effects			
Soft-tissue sarcoma	10/11 = 90.9%	88.10%	66.0%–99.1%
Salivary gland	10/12 = 83.3%	81.80%	58.0%–96.8%
IFS	7/7 = 100%	93.30%	70.5%–100%
Thyroid	5/5 = 100%	91.60%	63.0%–100%
Lung	3/4 = 75.0%	72.60%	30.4%–97.8%
Melanoma	2/4 = 50.0%	52.50%	12.4%–89.4%
Colon	1/4 = 25.0%	32.00%	2.6%–75.5%
GIST	3/3 = 100%	88.30%	49.3%–100%
Cholangiocarcinoma	0/2 = 0%	21.00%	0.0%–75.7%
Appendix	0/1 = 0%	30.00%	0.1%–89.7%
Breast	0/1 = 0%	30.00%	0.1%–90.1%
Pancreas	0/1 = 0%	29.80%	0.1%–89.7%
Unrepresented	—	56.90%	0.2%–99.9%

Reference: Murphy et al; Exploring Heterogeneity in Histology-Independent Technologies and the Implications for Cost-Effectiveness; Medical Decision Making 2021, Vol. 41(2) 165–178

4) Bayesian hierarchical models

How to estimate a subgroup you haven't observed?

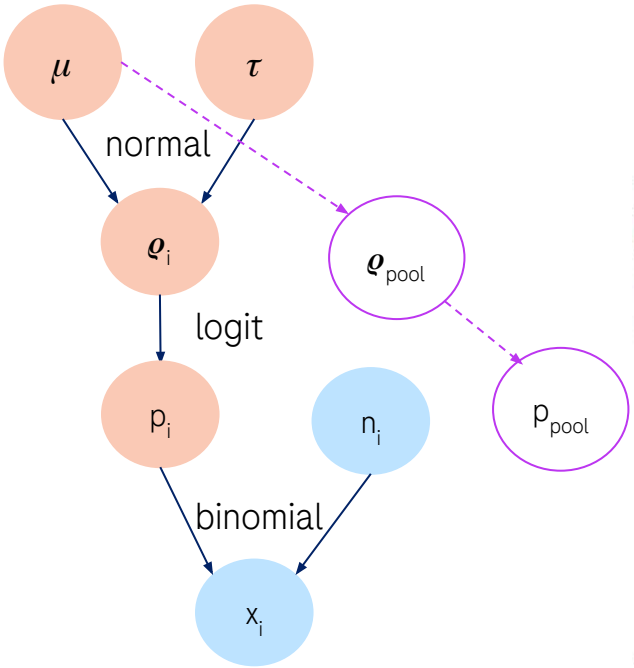


Table 2 Probabilities of response for all histologies

Histology	Observed Response	Estimated Mean Response Based on BHM (%)	95% CrI
Fixed effects			
Pooled	41/55 = 74.5%	74.20%	62.0%–84.7%
Random effects			
Soft-tissue sarcoma	10/11 = 90.9%	88.10%	66.0%–99.1%
Salivary gland	10/12 = 83.3%	81.80%	58.0%–96.8%
IFS	7/7 = 100%	93.30%	70.5%–100%
Thyroid	5/5 = 100%	91.60%	63.0%–100%
Lung	3/4 = 75.0%	72.60%	30.4%–97.8%
Melanoma	2/4 = 50.0%	52.50%	12.4%–89.4%
Colon	1/4 = 25.0%	32.00%	2.6%–75.5%
GIST	3/3 = 100%	88.30%	49.3%–100%
Cholangiocarcinoma	0/2 = 0%	21.00%	0.0%–75.7%
Appendix	0/1 = 0%	30.00%	0.1%–89.7%
Breast	0/1 = 0%	30.00%	0.1%–90.1%
Pancreas	0/1 = 0%	29.80%	0.1%–89.7%
Unrepresented	—	56.90%	0.2%–99.9%

Reference: Murphy et al; Exploring Heterogeneity in Histology-Independent Technologies and the Implications for Cost-Effectiveness; Medical Decision Making 2021, Vol. 41(2) 165–178

4) Bayesian hierarchical models

How to estimate a subgroup you haven't observed?

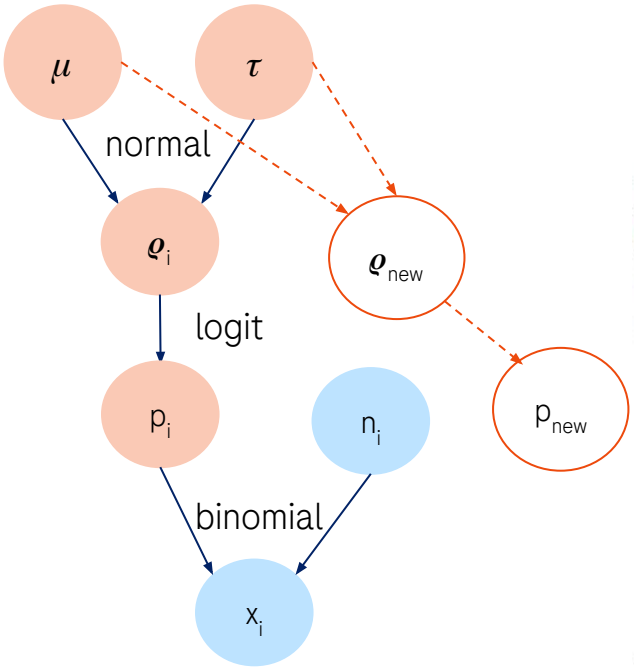


Table 2 Probabilities of response for all histologies

Histology	Observed Response	Estimated Mean Response Based on BHM (%)	95% CrI
Fixed effects			
Pooled	41/55 = 74.5%	74.20%	62.0%–84.7%
Random effects			
Soft-tissue sarcoma	10/11 = 90.9%	88.10%	66.0%–99.1%
Salivary gland	10/12 = 83.3%	81.80%	58.0%–96.8%
IFS	7/7 = 100%	93.30%	70.5%–100%
Thyroid	5/5 = 100%	91.60%	63.0%–100%
Lung	3/4 = 75.0%	72.60%	30.4%–97.8%
Melanoma	2/4 = 50.0%	52.50%	12.4%–89.4%
Colon	1/4 = 25.0%	32.00%	2.6%–75.5%
GIST	3/3 = 100%	88.30%	49.3%–100%
Cholangiocarcinoma	0/2 = 0%	21.00%	0.0%–75.7%
Appendix	0/1 = 0%	30.00%	0.1%–89.7%
Breast	0/1 = 0%	30.00%	0.1%–90.1%
Pancreas	0/1 = 0%	29.80%	0.1%–89.7%
Unrepresented	—	56.90%	0.2%–99.9%

Reference: Murphy et al; Exploring Heterogeneity in Histology-Independent Technologies and the Implications for Cost-Effectiveness; Medical Decision Making 2021, Vol. 41(2) 165–178

Were these approaches accepted?

Outcome is mixed

	Recommendation		Intrapatient included		RWD included	
	Larotrectinib	Entrectinib	Laro	Entrec	Laro	Entrec
England	+ (conditional)	+ (conditional)				
Germany	No additional benefit	No additional benefit				
France	+ (partial IFS, STS)	-				
Canada	- 2019 / + 2021	n/a				
Scotland	n/a	+				
Denmark	-	-				
Sweden	+ (partial <18)	+				

Reference: Brogaard, N., Abdul-Ghani, R., Bayle, A., Henderson, N., Bréant, A, and Steuten, L. 2021. Health technology assessment challenges associated with tumour-agnostic therapies: learnings from the assessments of entrectinib and larotrectinib. OHE Consulting Report, London: Office of Health Economics. https://www.ohe.org/publications/learnings-assessments-entrectinib-and-larotrectinib-health-technology-assessment#

Outcome is mixed, as was supplemental evidence included

	Recommendation		Intrapatient included		RWD included	
	Larotrectinib	Entrectinib	Laro	Entrec	Laro	Entrec
England	+ (conditional)	+ (conditional)	+	+	-	-
Germany	No additional benefit	No additional benefit	-	-	-	+
France	+ (partial IFS, STS)	-	-	-	-	-
Canada	- 2019 / + 2021	n/a	+ 2019 / + 2021	n/a	- 2019 / + 2021	n/a
Scotland	n/a	+	n/a	+	n/a	-
Denmark	-	-	+	+	-	-
Sweden	+ (partial <18)	+	+	-	-	+

Reference: Brogaard, N., Abdul-Ghani, R., Bayle, A., Henderson, N., Bréant, A, and Steuten, L. 2021. Health technology assessment challenges associated with tumour-agnostic therapies: learnings from the assessments of entrectinib and larotrectinib. OHE Consulting Report, London: Office of Health Economics. https://www.ohe.org/publications/learnings-assessments-entrectinib-and-larotrectinib-health-technology-assessment#

Influence of External control (real world data) varies

			Critique							
	Decision	**Influence RWD ECA**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
NICE	Recommended	did not review								
G-BA	No proof of added benefit	Low								
HAS	Do not recommend; ASMR N/A, SMR Insufficient	did not review								
pCODR	Recommended with restrictions	Low								
PBAC	Recommended	did not review								
	Sum across all four cases reviewed:									

Critiques

1) SoC inconsistent over time
2) ECA non-generalizable to clinical practice
3) Unmeasured confounding
4) Unjustified confounders
5) Selection bias
6) Incorrect adjusting
7) Inconsistent outcomes definitions
8) Data loss/insufficiency

Reference: Jaksa A, Louder A, Maksymiuk C, et al; A Comparison of Four Oncology External Control Arm Case Studies: Critiques From Regulatory and Health Technology Assessment Agencies; Value In Health 2022; in press; https://doi.org/10.1016/j.jval.2022.05.016

Influence of External control (real world data) varies

	Decision	Influence RWD ECA	Critique							
			1	2	3	4	5	6	7	8
NICE	Recommended	did not review								
G-BA	No proof of added benefit	Low		x	x					
HAS	Do not recommend; ASMR N/A, SMR Insufficient	did not review								
pCODR	Recommended with restrictions	Low			x		x			
PBAC	Recommended	did not review								
Sum across all four cases reviewed:			8	6	13	1	14	3	7	7

Critiques

1) SoC inconsistent over time
2) ECA non-generalizable to clinical practice
3) Unmeasured confounding
4) Unjustified confounders
5) Selection bias
6) Incorrect adjusting
7) Inconsistent outcomes definitions
8) Data loss/insufficiency

Reference: Jaksa A, Louder A, Maksymiuk C, et al; A Comparison of Four Oncology External Control Arm Case Studies: Critiques From Regulatory and Health Technology Assessment Agencies; Value In Health 2022; in press; https://doi.org/10.1016/j.jval.2022.05.016

Influence of External control (real world data) varies

	Decision	Influence RWD ECA	Critique 1	2	3	4	5	6	7	8
NICE	Recommended	did not review								
G-BA	No proof of added benefit	Low		x	x					
HAS	Do not recommend; ASMR N/A, SMR Insufficient	did not review								
pCODR	Recommended with restrictions	Low			x		x			
PBAC	Recommended	did not review								
Sum across all four cases reviewed:			8	6	**13**	1	**14**	3	7	7

Critiques

1) SoC inconsistent over time
2) ECA non-generalizable to clinical practice
3) Unmeasured confounding
4) Unjustified confounders
5) Selection bias
6) Incorrect adjusting
7) Inconsistent outcomes definitions
8) Data loss/insufficiency

Reference: Jaksa A, Louder A, Maksymiuk C, et al; A Comparison of Four Oncology External Control Arm Case Studies: Critiques From Regulatory and Health Technology Assessment Agencies; Value In Health 2022; in press; https://doi.org/10.1016/j.jval.2022.05.016

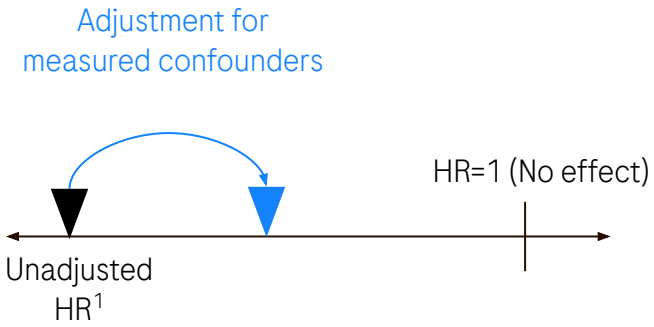
What can we do better?

Unmeasured confounders - quantitative bias analysis

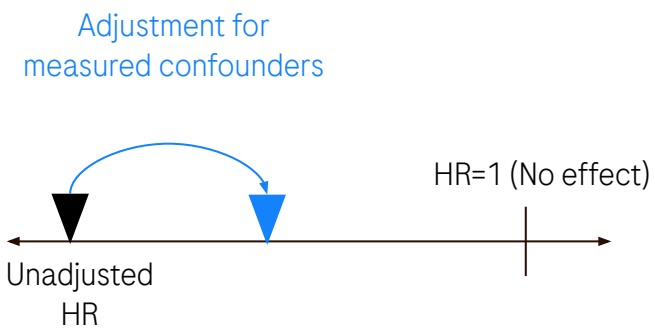
There are known unknowns and unknown unknowns

- Two methodologies represented below:
  - Using external information to correct for bias (“**external adjustment**”)
  - **Tipping point analysis**

External adjustment



Tipping point analysis



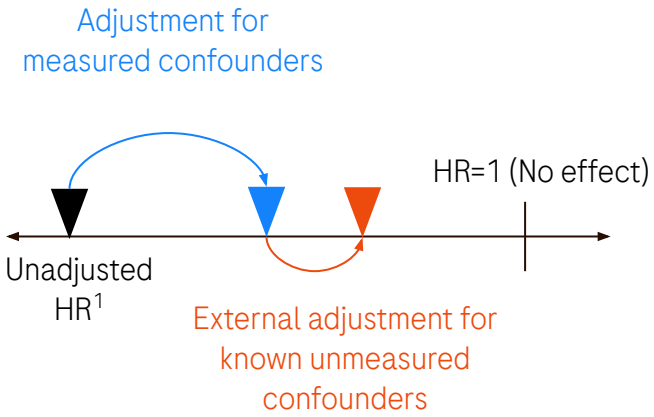
[1]HR = hazard ratio

Unmeasured confounders - quantitative bias analysis

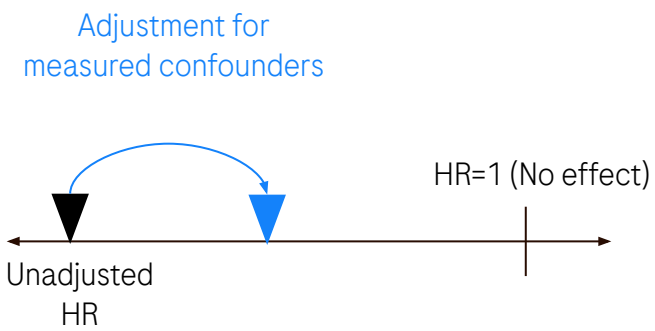
There are known unknowns and unknown unknowns

- Two methodologies represented below:
  - Using external information to correct for bias (“**external adjustment**”)
  - **Tipping point analysis**

External adjustment



Tipping point analysis



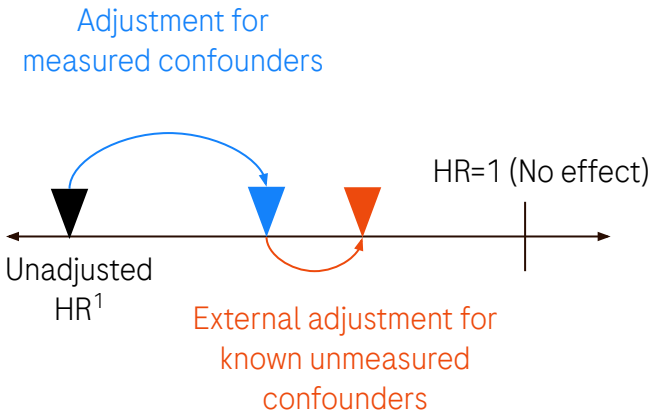
[1]HR = hazard ratio

Unmeasured confounders - quantitative bias analysis

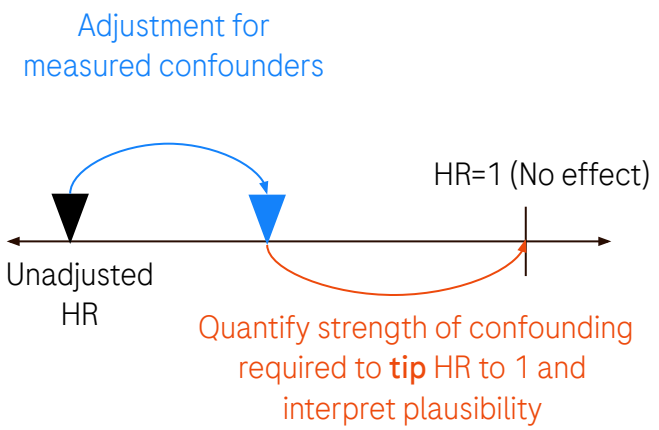
There are known unknowns and unknown unknowns

- Two methodologies represented below:
  - Using external information to correct for bias (“**external adjustment**”)
  - **Tipping point analysis**

External adjustment



Tipping point analysis



[1]HR = hazard ratio

Example: tipping point - known unknown

ECOG is known confounder missing in RWD

Negative values of δ imply exponentially increasing odds of patients having poorer ECOG PS than expected under missing at random (MAR) given their covariates.

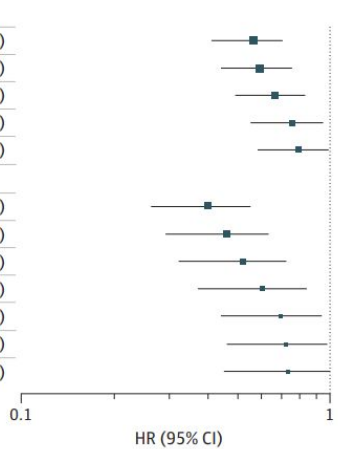
Figure 2. Tipping Point Analysis for Missing Eastern Cooperative Oncology Group (ECOG) Performance Status (PS)

δ=0 (MAR)

poorer ECOG than MAR

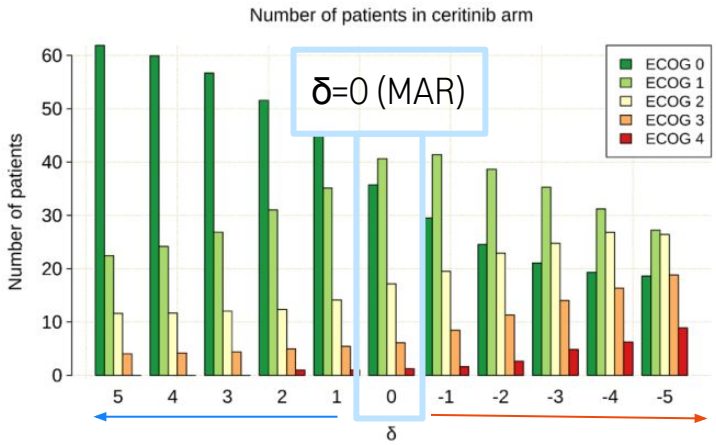
δ=0 (MAR)

poorer ECOG than MAR



HR: Alectinib vs Ceritinib

Distribution of ECOG PS in ceritinib RWD cohort



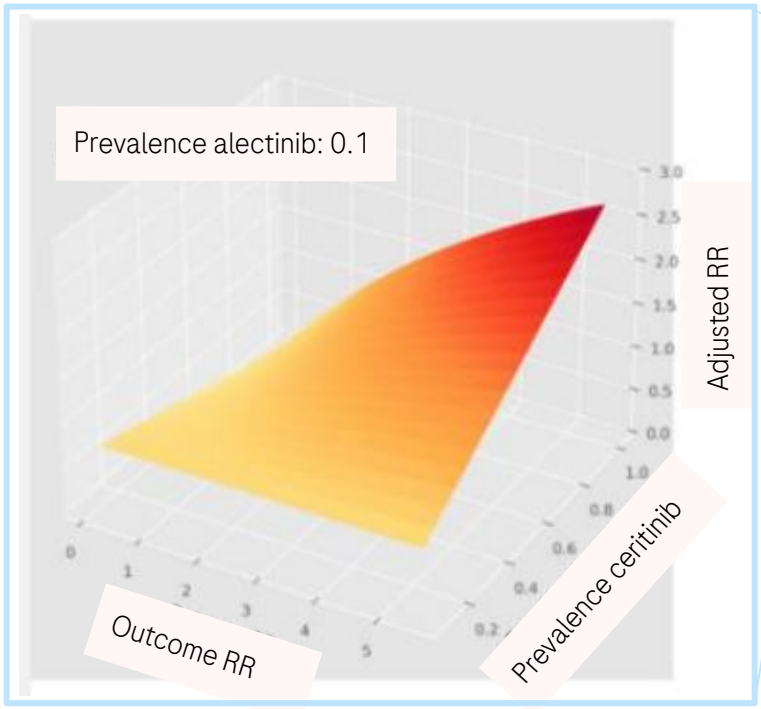
Missing data better ECOG than MAR

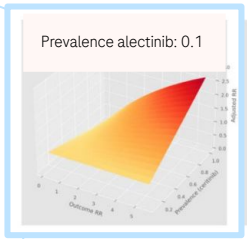
Missing data poorer ECOG than MAR

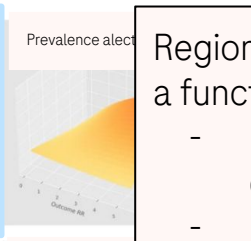
Reference: Wilkinson et al; Assessment of Alectinib vs Ceritinib in ALK-Positive NSCLC in Phase 2 Trials and Real-world Data; JAMA Network Open. 2021;4(10):e2126306. doi:10.1001/jamanetworkopen.2021.26306

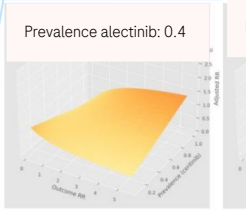
Example: tipping point - unknown unknown

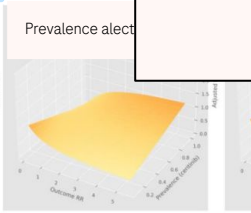
How large would effect of an unmeasured confounder need to be to change conclusions?

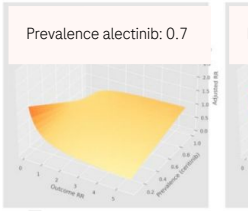


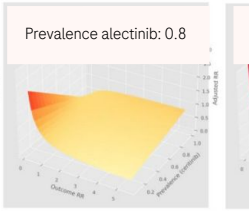


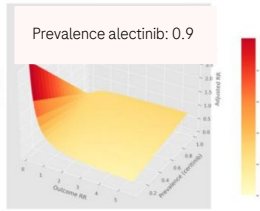


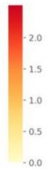












Region where conclusions change is a function of:

- Effect of the unknown confounder on outcome
- Imbalance in that confounder prevalence between arms

Dark red values represent regions where adjusted RR would “tip” base case conclusions

Reference: Wilkinson et al; Assessment of Alectinib vs Ceritinib in ALK-Positive NSCLC in Phase 2 Trials and Real-world Data; JAMA Network Open. 2021;4(10):e2126306. doi:10.1001/jamanetworkopen.2021.26306

Conclusions

Conclusions

- Single arm trials will continue to be a core piece of evidence packages
- Statistical methods and research approaches exist to complement these and explore uncertainties
- HTA bodies are making different decisions in different contexts with different concerns

How can we together better develop and present evidence to support decision making under uncertainty?

Doing now what patients need next