Indirect comparisons w/o adjustment for patient characteristics within the framework of AMNOG

Sarah Böhme

PSI One Day Meeting and Workshop: Real World Evidence Generalisability of Treatment Comparisons for Decision Making

Bad Homburg, 18 September 2018

Disclaimer

The opinions, thoughts and viewpoints expressed in this talk are those of the author and do not necessarily reflect the opinions, thoughts and viewpoints of Pfizer.

Outline

Background

- for situations where the appropriate comparator as determined by Federal Joint Committee (G-BA) is not met by the comparator in a clinical trial AMNOG allows indirect comparisons (IC) to assess extent of added benefit
- IQWiG generally only accepts adjusted IC
- only if a dramatic effect can be demonstrated an unadjusted IC might be accepted

- Werner et al. (2017) conducted an evaluation of IC assessed by IQWiG between 2011 and Oktober 2017
  - 94% of $n = 111$ presented indirect comparisons were disapproved by IQWIG
  - of those, 25% have not been accepted due to inapproriate statistical method and 40% due to unsimilarity of the studies in the IC

Indirect comparisons (IC)

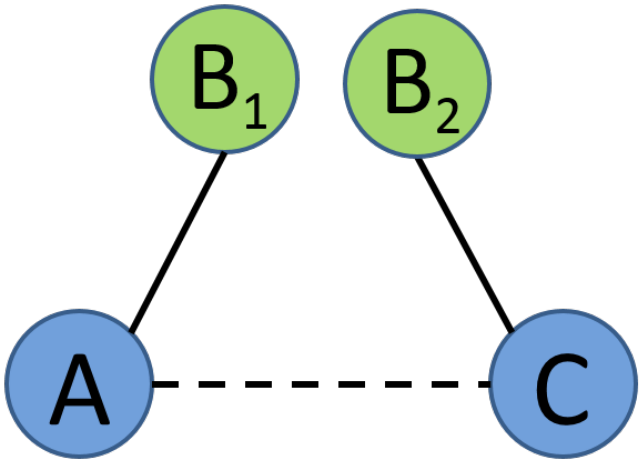
Adjusted IC

- e.g. Bucher approach
- maintain randomization
- several requirements, e.g. connected network and similarity of trials to be compared
- biased if there are imbalances in effect modifiers between trials
- other methods using network meta-analysis: Bayes (MTC), Lumley

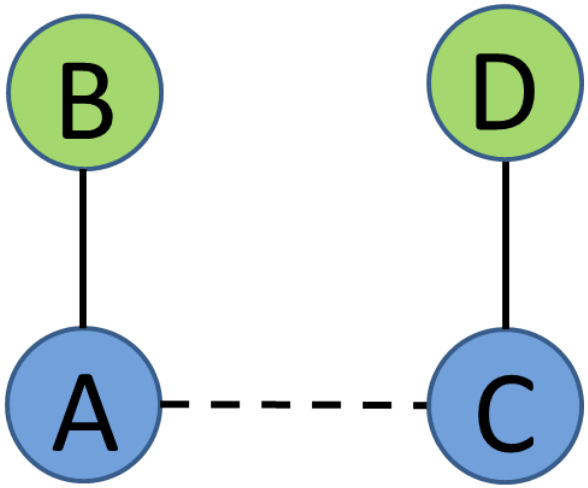
Unadjusted IC

- e.g. historical, naive, Matching-Adjusted IC
- break randomization
- always potential risk of (substantial) bias
- do not necessarily require a connected network

Matching Adjusted Indirect Comparison (MAIC)



- able to overcome issue of unbalanced study populations



- able to overcome issue of a disconnected network

MAIC- Signorovitch et al., 2010

- population adjustment approach
- uses individual patient data (IPD) for one treatment (index trial) and aggregated data for the other treatment of interest
- seeks to adjust for imbalances in effect modifiers and derive adjusted outcomes

Assumptions:

- general comparibility of studies in terms of
  - type of population included (e.g. not 1st line vs. 2nd line)
  - measurement methods
  - time frame of observation
  - reporting of information
  - ...
- complete information on relevant effect modifiers from both trials
- effect modifiers influence outcome in same way in both trials
- study populations must have sufficient overlap

MAIC - Signorovitch et al., 2010

Concept:

- reweight patients in IPD trial so that weighted average values of determinants of outcomes in the index trial match aggregated data
- weights derived from propensity score-type analysis using IPD

$$w_i = \frac{P(T_i{}^1 = 1|x_i)}{P(T_i = 0|x_i)}$$

- individual weights applied in analyses (e.g. weighted means, weighted Cox regression) to generate adjusted outcome and confidence interval

[1] $T_i \hat{=}$ treatment/study index

Objective

- adjusted IC (e.g. Bucher approach) gold standard, but biased in case of imbalances in effect modifiers between trials and not applicable in case of disconnected networks
- MAIC seeks to adjust for such imbalances and is applicable in case of disconnected networks

- aim to evaluate properties of Matching-Adjusted Indirect Comparison for the analysis of survival endpoints in various scenarios
  1. common comparator present, but imbalance in an effect modifier between study populations
  2. lack of common comparator $\Rightarrow$ disconnected network

Simulate IPD

- generation of IPD from H2H trials (A vs. B and C vs. B)
- outcome: survival endpoint
- generation of individual Weibull distributed survival times to simulate Cox proportional hazards models
- constant treatment effect in comparator trial of $HR_{CB} = 0.85$
- simulated treatment effects for indirect comparison HR_{AC}:
  - major effect $\equiv$ upper CI < 0.85 $\Rightarrow HR_{AC} = 0.50$
  - considerable effect $\equiv$ upper CI < 0.95 $\Rightarrow HR_{AC} = 0.83$
  - minor effect $\equiv$ upper CI < 1.00 $\Rightarrow HR_{AC} = 0.95$
  - no treatment effect $\equiv$ $1.00 \in$ CI $\Rightarrow HR_{AC} = 1.00$

Further simulation design

- sample size calculation for H2H trials
  - HR_{AB} or HR_{CB}
  - $\alpha = 0.05$, Power $= 0.8$
  - no censoring applied
- doubling of derived sample size

- **Settings:**

  1. common comparator, but imbalanced study populations with regard to effect modifier X and interaction effect between T and X
     a. moderate imbalance
     b. strong imbalance
  2. lack of common comparator, $\nu = 1$, $\nu = 0.8$, $\nu = 1.8$
     - balanced study populations

- run 10,000 iterations each

Results - (1a) moderate imbalanced study populations

Table 1: (1a) Estimated treatment effects derived by IC, Power and Type I error - interaction effect $\log(0.97)$

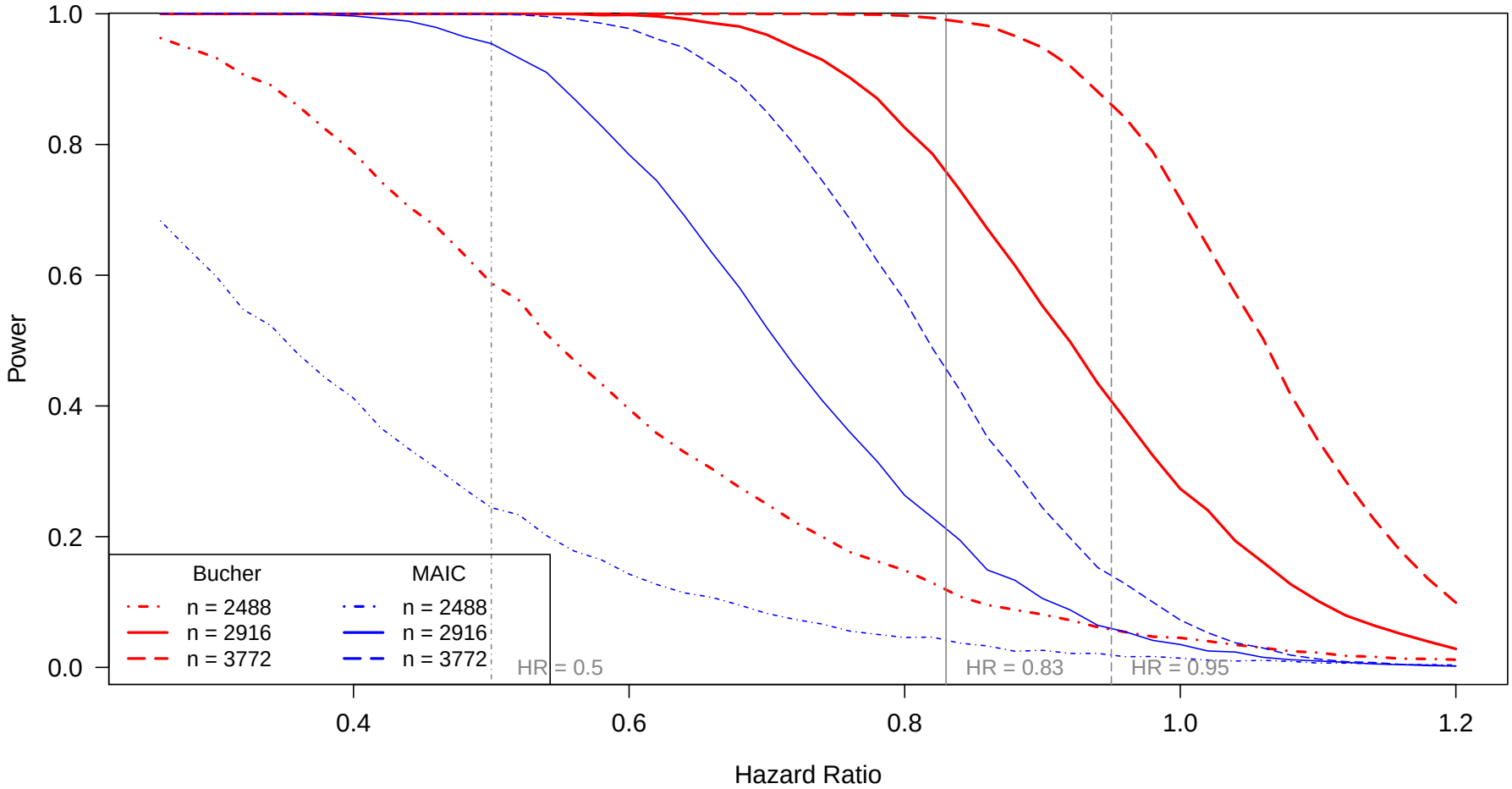
Effect	True HR	Bucher		MAIC	
		$\widehat{HR}$	Power/Type I error	$\widehat{HR}$	Power/Type I error
major	0.50	0.403	0.544	0.490	0.283
considerable	0.83	0.653	0.753	0.822	0.135
minor	0.95	0.743	0.877	0.944	0.083
no	1.00	0.782	0.859	0.994	0.059

Table 2: (1a) Estimated treatment effects derived by IC, Power and Type I error - interaction effect $\log(1.05)$

Effect	True HR	Bucher		MAIC	
		$\widehat{HR}$	Power/Type I error	$\widehat{HR}$	Power/Type I error
major	0.50	0.792	0.074	0.648	0.228
considerable	0.83	1.254	< 0.001	0.910	0.068
minor	0.95	1.427	0.000	1.002	0.035
no	1.00	1.502	0.980	1.055	0.083

Results - (1a) moderate imbalanced study populations

Power functions



Results - (1b) strong imbalanced study populations

Table 3: (1b) Estimated treatment effects derived by IC, Power and Type I error - interaction effect $\log(0.97)$

Effect	True HR	Bucher		MAIC	
		$\widehat{\text{HR}}$	Power/Type I error	$\widehat{\text{HR}}$	Power/Type I error
major	0.50	0.365	0.679	0.486	0.306
considerable	0.83	0.590	0.932	0.813	0.126
minor	0.95	0.672	0.990	0.944	0.088
no	1.00	0.707	0.992	0.991	0.072

Results setting (1) - summary & possible explanations

- Bucher clearly biased and not appropriate
- MAIC:
  - able to adjust outcome, however still (slightly) biased in 2 of the 3 settings
  - type I error slightly above the nominal ($\alpha = 0.05$)
  - moderate imbalance results in reduction of effective sample size by $1/3$, strong imbalance by $2/3$ for adjusted population
- reasonable explanation: only adjustment for balanced population mean but not with regard to standard deviation (2nd moment)
  $\Rightarrow$ should be controlled and if not balanced further adjustment might be needed
  $\Rightarrow$ would reduce effective sample size even more
- power of indirect comparisons generally low compared to direct comparison (see also Kuehnast et al., 2017)

MAIC - survival endpoints and disconnected networks

Problem:

- adjustment of outcome trivial in case of continuous endpoints
- in case of survival endpoints adjustment of outcome nontrivial as weights cannot just be applied on individual outcomes
- adjusted HR can be derived by using a weighted Cox regression model, but requires common comparator

Idea:

- using ratio of median survival time estimates (Median Ratio) as a surrogate measure for HR, where median from IPD is derived by weighted Kaplan-Meier
- as the use of the "plain" Median Ratio requires exponential distributed suvival times, we tried to "adjust" Median Ratio by exponentiating with estimated shape parameter $\hat{\nu}$ (Ishak et al., 2015)

Results - (2) balanced study populations

Table 4: (I) Estimated treatment effects derived by MAIC, Power and Type I error

Effect	True HR	Estimated Median ratio $\nu = 1$	$\nu = 0.8$	$\nu = 1.8$	Power/Type I error $\nu = 1$	$\nu = 0.8$	$\nu = 1.8$
major	0.50	—— 0.605 ——			0.750	0.815	0.421
considerable	0.83	—— 0.878 ——			0.409	0.505	0.120
minor	0.95	—— 0.961 ——			0.312	0.398	0.080
no	1.00	—— 0.998 ——			0.230	0.336	0.042

Results setting (2) - summary

- our approach using adjusted Median Ratio as surrogate measure for HR yielded (slightly) biased results
- applied variance estimator for Median Ratio does not seem to be appropriate, as type I errors were partially far too high and differed strongly between different values of shaping parameter ν
- adjusted Median Ratio as applied in the simulation study not appropriate

Conclusion I

- IC important option in case of lack of direct evidence
- adjusted IC reflect gold standard but require a common comparator and balanced study populations, which is often not the case
- MAIC is a population adjustment approach to conduct an (unadjusted) IC, which can be applied in situations of imbalanced study populations
- MAIC requires information on all relevant effect modifiers to be able to adjust for them and cannot adjust for unobserved effect modifiers $\Rightarrow$ should be applied with care
- balance of at least 1st and 2nd moment of effect modifier should be checked after reweighting of patients
- effective sample sizes and distributions of weights should be checked

Conclusion II

- application of MAIC on continous endpoints straightforward and already often described in the literature regardless of presence or absence of a common comparator
- for disconnected networks application on survival endpoints *not straightforward* and needs to be further investigated, however likely to be (substantially) biased
- even in case of straightforward outcome adjustment, an MAIC assumes that all effect modifiers and prognostic factors are accounted for, which seems hardley to be meet in practice
- evidence needed that MAIC performs any better than naive approaches
- further research on MAIC approach necessary
- *useful source*: NICE DSU Technical Support Document 18 (2016)

References I

Bender R, Augustin T and Blettner M. *Generating Survival Times to Simulate Cox Proportional Hazards Models.* German Research Foundation (DFG); Sonderforschungsbereich 386 2003; Paper 338.

Bucher HC, Guyatt GH, Griffith LE, Walter SD. *The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials.* Journal of Clinical Epidemiology. 1997; 50(6): 683-691.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). *Allgemeine Methoden Version 5.0 vom 10.07.2017.* 2017. Available from: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf.

Ishak KJ, Proskorovsky I and Benedict A. *Simulation and Matching-Based Approaches for Indirect Comparison of Treatments.* Pharmacoeconomics 2015; 33: 537-549.

Kuehnast S, Schiffner-Rohe J, Rahnenführer J and Leverkus F. *Evaluation of Adjusted and Unadjusted Indirect Comparison Methods in Benefit Assessment - A Simulation Study for Time-to-event Endpoints.* Methods Inf Med 2017; 56: 261-267.

Michiels S, Piedbois P, Burdett S, Syz N, Stewart L and Pignon JP. *Meta-analysis when only the median survival times are known: A comparison with individual patient data results.* International Journal of Technology Assessment in Health Care 2005; 21(1): 119-125.

References II

Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR and Welton NJ. *NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE.* 2016. Available from http://www.nicedsu.org.uk

Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR and Welton NJ. *Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal.* Medical Decision Making 2018; 38(2) 200-211.

Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, Gupta SR and Mulani PM. *Comparative Effectiveness Without Head-to-Head Trials: A Method for Matching-Adjusted Indirect Comparisons Applied to Psoriasis Treatment with Adalimumab or Etanercept.* Pharmacoeconomics 2010; 28(10): 935-945.

Signorovitch JE, Swallow E, Kantor E, Wang X, Klimovsky J, Haas T, Devine B and Metrakos P. *Everolimus and sunitinib for advanced pancreatic neuroendocrine tumors: a matching-adjusted indirect comparison.* Experimental Hematology & Oncology 2013; 2: 32.

Skipka G, Wieseler B, Kaiser T, Thomas S, Bender R, Windeler J and Lange S. *Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs.* Biometrical Journal 2016; 58(1): 43-58.

Werner S, Lechterbeck L and Rasch A. *Akzeptanz und Ablehnungsgründe von indirekten Vergleichen in IQWiG Nutzenbewertungen.* In Preparation 2017.

Back up

Simulation of Weibull distributed survival times T[1]

$$T = \left(\frac{-log(u)}{h_0 \exp(X\beta)} \right)^{\frac{1}{\nu}}$$

- $u \sim \mathcal{U}[0, 1]$
- design matrix X
- $\beta = (\beta_1, ..., \beta_p)$
- scale parameter $\lambda = h_0 \exp(X\beta)$
- shape parameter ν

[1] Bender et al., 2003

MAIC - adjusted Median Ratio

$$MR_{AC}^{adj} = \left(\frac{M_C}{M_A^*} \right)^{\hat{\nu}}$$

$$\text{Var}(\log MR_{AC}^{adj}) = \hat{\nu}^2 \left(\frac{1}{d_A} + \frac{1}{d_C} \right)$$

- MR_{AC}^{adj}: adjusted ratio of median survial times
- M_A^*: weighted median survival time in treatment arm A
- M_C: median survival time in treatment arm C
- $\hat{\nu}$: shape parameter estimate
- d_A, d_C: number of events in treatment arm A and C, respectively