IQWiG's methodological approach to subgroup analyses in benefit assessments

Ulrich Grouven, Guido Skipka

IQWiG

Agenda

- What is IQWiG?
- Subgroup analyses in IQWiG assessments
  - Legal requirements
  - Methods
  - Historical development
- An illustrating example
- Outlook
- Summary
- References

What is IQWiG (Institute for Quality and Efficiency in Health Care) ?

- independent scientific institute
- founded in 2004 by Federal Joint Committee (G-BA: highest decision-making body of the joint self-government of physicians, dentists, hospitals and health insurance funds in Germany)
- legal foundations and responsibilities laid down in the German “Social Code Book V (SGB V) - Statutory Health Insurance”
- commissioned by G-BA or the Federal Ministry of Health (BMG)

Tasks

- examines the benefits and harms of medical interventions (i.e. new drugs, surgical procedures, diagnostic procedures) for patients and other affected persons
- provides information on the advantages and disadvantages of these interventions (for decision makers, experts and also for lay people)

IQWiG's principles

- evidence-based: specified in IQWiG's General Methods
- independent: no influence on content of reports by payers, service providers, industry organizations or politicians
- patient-orientated: assessment of patient-relevant outcomes, involvement of patients and other affected persons
- transparent: publication of all documents relevant for reports and of the methods paper; disclosure of conflicts of interest by all persons involved in reports (employees, external experts etc.)

→ IQWiG provides the scientific basis for the G-BA's decisions
(e. g. introduction of health services, reimbursement of new drugs)

Subgroup analyses

Important aspect in IQWiG's evaluations

- especially relevant in *dossier assessments*
  - question: has a new drug an added benefit compared with standard therapy?
  - pharmaceutical company submits dossier to demonstrate the benefits of the new drug
  - results relevant for reimbursement decisions

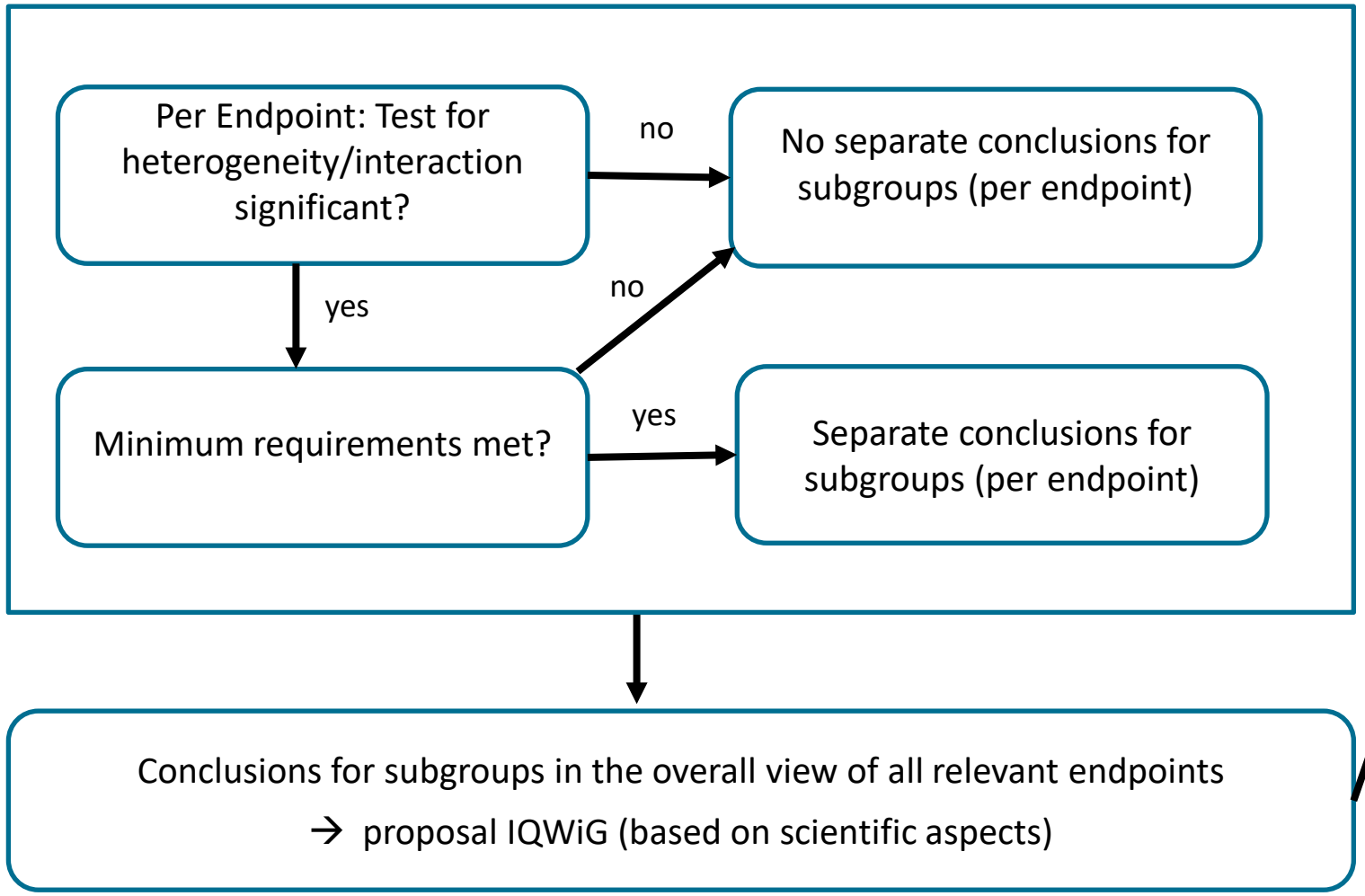
- Aims
  - examination whether the results (treatment effects) differ between different subgroups and identification of patient subgroups with varying benefit / harm of the intervention
  - leads to potentially different added benefit for different patient groups

Statutory requirements

- According to §139a (2) SGB V, the Institute is obliged to consider characteristics specific to age, gender, and life circumstances. In addition, it should also be elaborated in which patient groups a new drug is expected to lead to a relevant improvement in treatment success, with the aim of providing these patients with access to this new drug [170]. A corresponding objective can also be found in §35a SGB V regarding the assessment of the benefit of drugs with new active ingredients [171]. In this assessment, patient groups should be identified in whom these drugs show a therapeutically relevant added benefit.

- 170. Deutscher Bundestag. Gesetzentwurf der Fraktionen SPD, CDU/CSU und BÜNDNIS 90/DIE GRÜNEN: Entwurf eines Gesetzes zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG) [online]. 2003 [Accessed: 04.02.2021]. URL: http://dipbt.bundestag.de/doc/btd/15/015/1501525.pdf.
- 171. Deutscher Bundestag. Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz – AMNOG) vom 22. Dezember 2010. Bundesgesetzblatt Teil 1 2010; (67): 2262-2277.

Methodological approach (I)



G-BA:
→ final decision

based on:

- IQWiG assessment
- Comments and arguments from stakeholder / interested parties
- Possible other aspects

Methodological approach (II)

- first step:

  → test for heterogeneity / interaction

- use of standard methods*:

  → Cochran‘s Q-Test

- standard significance level:

  → $\alpha = 0,05$

* in case own calculations are necessary

Methodological approach (III)

Additional criteria (minimal requirements)

- each subgroup comprises at least 10 people
- at least 10 events occurred in one of the subgroups (in case of binary data and survival times)
- significant and relevant effect in at least 1 subgroup

→ pragmatic rules to ensure sufficient reliability of the results

Short historical review (I)

Subgroup analyses (previous procedure)

Test for interaction between subgroup factor and treatment

- p-value < 0.05
  - *proof* for different subgroup effects
  - separate statements for subgroups without consideration of the overall result
- 0.05 ≤ p-value < 0.20
  - *indication* of different subgroup effects
  - separate statements for subgroups taking into account the overall result

Disadvantages: high resource requirements and complexity

Short historical review (II)

Empirical investigation

In dossier assessments in 2015, 100 subgroup analyses (for 22 research questions) were analysed and revealed 3 cases with indications of effect modification (i.e. $0{,}05 \leq p < 0{,}20$), of which only 1 indication was considered in the final G-BA‘s decision.

Consequence

→ Introduction of simplified procedure for subgroup analyses (as described) since General Methods 5.0 (2017)

→ Advantages

- reducing resource requirements and complexity
- reducing the problem of multiplicity
- welcomed and approved by the decision-makers

Remarks

- **More than 2 subgroups**
  - pairwise statistical tests
  - not statistically significant pairs ($\alpha = 0.05$) summarized into one group (if meaningful)

- **Effect modification for > 1 subgroup characteristic (interaction of higher order)**
  - interpretation difficult
  - separate analyses required for combined subgroups
  - such analyses are rarely available
  - decision on a case-by-case basis

Example

Dossier assessment: [A22-137] Abemaciclib

- **Indication**

  Postmenopausal women with hormone receptor-positive, HER2-negative locally advanced or metastatic **breast cancer** who have not yet received initial endocrine-based therapy

- **Intervention**

  Abemaciclib (combination with an aromatase inhibitor)

- **Control**

  Appropriate comparator therapy (ACT) (current standard therapy, determined by the G-BA): Anastrozole or Letrozole (or …)

Example (description)

- **Relevant outcomes**

  overall survival, PROs (EORTC-C30 score for morbidity / HRQoL), adverse events

- **2 relevant studies (RCTs, parallel, double-blind)**

  MONARCH 3 (n=493), MONARCH plus (n=396)

- **Subgroup analyses**
  - gender (→ not applicable)
  - age (< 65 years, ≥ 65 years)
  - type of disease (visceral metastases versus non-visceral metastases)

Example (results I)

Subgroup analyses (age)

Exemplary:

results for serious adverse events (SAE)

Outcome Characteristic Study Subgroup	Abemaciclib + anastrozole or letrozole		Placebo + anastrozole or letrozole		Abemaciclib + anastrozole or letrozole versus placebo + anastrozole or letrozole	
	N	Median time to event in months [95% CI] Patients with event n (%)	N	Median time to event in months [95% CI] Patients with event n (%)	HR [95% CI][a]	p-value[b]
SAEs						
MONARCH 3	327	50.3 [38.4; 65.9] 122 (37.3)	161	NR 29 (18.0)	1.95 [1.30; 2.93]	0.001
MONARCH plus	205	NR 56 (27.3)	99	NR 11 (11.1)	2.17 [1.13; 4.14]	0.016
Total					2.01 [1.42; 2.83]	< 0.001
Age						
MONARCH 3						
< 65 years	179	65.9 [46.45; NC] 49 (27.4)	89	NR [36.56; NC] 16 (18.0)	1.15 [0.65; 2.04]	0.622
≥ 65 years	148	27.2 [14.27; 48.33] 73 (49.3)	72	NR 13 (18.1)	3.17 [1.76; 5.73]	< 0.001
MONARCH plus						
< 65 years	156	NR 34 (21.8)	83	NR 7 (8.4)	1.92 [0.85; 4.36]	0.112
≥ 65 years	49	20.4 [7.63; NC] 22 (44.9)	16	NR [17.06; NC] 4 (25.0)	2.70 [0.93; 7.88]	0.058
Total					Interaction:	0.023
< 65 years					1.36 [0.85; 2.17]	0.196
≥ 65 years					3.06 [1.82; 5.12]	< 0.001

a. HR: unstratified Cox model; meta-analysis: fixed effect model.
b. p-value: log-rank test; interaction p-value from metaanalysis (Cochran's Q-test for heterogeneity).

Excursion

How does IQWiG rate the available evidence?

2 relevant aspects

1. Certainty of results
   - → based on number and quality of available study results
   - → categories: proof / indication / hint
   - → *qualitative* assessment
2. Extent of added benefit
   - → based on size of effect and type of outcome
   - → categories: major / considerable / minor / not quantifiable
   - → *quantitative* assessment

Example (results II)

Favourable:

- overall survival
  (HR: 0,78 [0,63; 0,98]; 0,034)

(→ proof of minor added benefit)

Unfavourable
(varying certainty and extent):

- various items of EORTC (morbidity and HRQoL)
- adverse events

Subgroup specific results for

- serious adverse events
- Global Health status (EORTC)
- Social functioning (EORTC)

(for age ≥ 65 years)

IQWiG assessment categories	
Certainty of results	Proof / indication / hint
Extent of added benefit	Major / considerable / minor / not quantifiable

Example (results III)

- **Overall conclusion (IQWiG)**

  → no added benefit

  (“In summary, weighing the favourable effect of minor extent against the numerous unfavourable effects of at most major extent. … thus, there is no proof of added benefit of abemaciclib + anastrozole or letrozole compared with anastrozole or letrozole.”)

  → no different conclusion regarding different subgroups

Example (results IV)

- **Final conclusion (G-BA)**

  → hint for minor added benefit

  (“In a weighing-up decision, the G-BA comes to the conclusion that the improvement in overall survival outweighs the significant disadvantages in terms of side effects and other disadvantages with regard to the symptoms of the disease.”)

  → no different conclusion regarding different subgroups

  (“Overall, the significance of the available subgroup results for the assessment of added benefit is considered insufficient.”)

Outlook (alternative approaches) (I)

1 subgroup per study

- *Generalized Q-Tests* proposed by several authors
  - use different weights
  - provide a more accurate estimation of the amount of heterogeneity

- *F-test* (based on meta-regression)
  - own work in progress: simulation study comparing Q-test and F-test for subgroup analyses (especially with few studies)
    - → lower type-1-error for F-test (significance level was maintained)
    - → closer investigation of power still pending

Outlook (alternative approaches) (II)

>1 subgroup per study

„Subgroup first approach “

- Generalized Q-Tests

„Study first approach “

- Within-trial framework (Godolphin et al. 2022)

  Stepwise procedure:

  (1) estimating within-trial interactions across two or more subgroups

  (2) estimating subgroup-specific treatment effects ("floating" estimates”) by using pooled interaction effects

Outlook (alternative approaches) (III)

Mixture of subgroup-first and study-first approach

→ see poster of Frank Weber (IQWiG):

„Meta-analyses with subgroups of patients: From subgroup-first to trial-first and stopping in between”

Summary

- Subgroup analyses are an important part of IQWiG's assessments
- When conducting subgroup analyses, IQWIG must comply with legal requirements
- IQWiG has developed standard procedures for conducting subgroup analyses, which are
  - efficient
  - pragmatic (tight timeframes and strict deadlines)
  - flexible
  - transparent
  - open for new developments
- IQWiG follows current developments and conducts its own research in order to continuously improve its methods

References

- IQWiG General Methods Version 7.0 (https://www.iqwig.de/en/about-us/methods/methods-paper/)
- Bowden J, et al. Quantifying, displaying and accounting for heterogeneity in the meta-analysis of RCTs using standard and generalised Q statistics. BMC Medical Research Methodology 2011; 11:41.
- Godolphin PJ et al. Estimating interactions and subgroup-specific treatment effects in meta-analysis without aggregation bias: A within-trial framework. Res Synth Methods 2023; 14(1):68-78.
- Guddat C. An investigation of the type I error rate when testing for subgroup differences in the context of random-effects meta-analyses. In: Filtering the information overload for better decisions. Abstracts of the 23rd Cochrane Colloquium; 2015 3-7 Oct; Vienna, Austria. John Wiley & Sons; 2015
- Kulinskawa E, Hoaglin DC. On the Q statistic with constant weights in meta-analysis of binary outcomes. BMC Medical Research Methodology 2023; 23:146.
- van Aert RCM, et al. Statistical properties of methods based on the Q-statistic for constructing a confidence interval for the between-study variance in meta-analysis. Res Syn Meth. 2019; 10:225–239.

Institute for Quality and Efficiency in Health Care (IQWiG

Thank you for your attention!

Questions and comments:

PD Dr. Ulrich Grouven

ulrich.grouven@iqwig.de

Im Mediapark 8
50670 Köln

Telefon +49 221 35685-0
Telefax +49 221 35685-1

info@iqwig.de

www.iqwig.de
www.gesundheitsinformation.de
www.themencheck-medizin.de

@iqwig@wisskomm.social
@iqwig_gi@wisskomm.social