Assessing Heterogeneity Across Groups of Clinical Trials Based on Aggregate Data

H Schmidt, H Köglé, S Geier, T Glaab, N Metzdorf
Boehringer Ingelheim Pharma GmbH & Co. KG, Ingegheim am Rhein and Biberach an der Riss, Germany; e-mail: hendrik.schmidt@boehringer-ingelheim.com

BACKGROUND

Meta-analyses are used to synthesize treatment information through application of statistical methods during all phases of clinical research and development, especially in the context of health technology assessment of the relative effectiveness between two treatments across trials as well as in drug reimbursement applications.1

The consistency of treatment effect across trials can be examined using different tests (eg, Cochran’s Q-test) or measures (eg, the I^2 measure). If trials included in meta-analysis are grouped according to certain criteria, heterogeneity between trial groups can also be assessed by applying a test for subgroup difference.

AIMS

To show that heterogeneity of treatment effects across groups of trials can be investigated using meta-analytic methods for aggregate data, as illustrated by two examples stemming from chronic obstructive pulmonary disease (COPD).

To show how meta-analysis software can be employed and has been applied in this context.

METHODS

Data Sources

Meta-analyses summarize relative treatment effects comparing two therapies across several trials based on aggregate trial information (relative treatment effect and its variability).

Heterogeneity of treatment effect across trials can be examined using different tests (eg, Cochran’s Q-test) or measures (eg, the I^2 measure). If trials included in meta-analysis are grouped according to certain criteria, heterogeneity between trial groups can also be assessed by applying a test for subgroup difference.

This is demonstrated using two examples from COPD.

Example 1 (exacerbations analysis)

The best of tiotropium was assessed by the German Institute for Quality and Efficiency in Healthcare (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen [IQWiG]).2

Heterogeneity was examined across controlled trials comparing different long-acting β2-agonists (LABAs; indacaterol, salmeterol and formoterol) with regard to their effectiveness for preventing exacerbations in COPD was included in the analysis.

The analysis included 19 randomized, controlled trials (indacaterol: n=2; salmeterol: n=3; formoterol: n=14).

Example 2 (quality of life [QoL] analysis)

In an analysis by Schmidt et al., it was examined whether the type of trial design (open label or double-blinded) was a potential modifier of treatment effect for health-related QoL, for a group of 15 randomized, placebo-controlled trials investigating the use of inhaled tiotropium (COPD).

QoL outcomes were assessed using the Saint George’s Respiratory Questionnaire (SGRQ), including SGRQ total score and SGRQ nonresponder rates.

Statistical Analysis

Classical meta-analysis

Classical meta-analyses comparing two treatments across several trials was applied.

From each trial, a measure of relative treatment effect was used along with a measure of its variability (95% confidence interval).

Trials were grouped and meta-analysis was performed within and across groups.

Heterogeneity between trial groupings was assessed via a Chi-Square (χ^2) test where the distribution of the test statistic is asymptotically χ^2-distributed with one degree of freedom.

For analysis, the RevMan software was applied. Additionally, a Boehringer Ingelheim internal meta-analysis software was used for recalculation and confirmation of results.

Example 1 (exacerbations analysis)

The clinical endpoint used for analysis was the number of patients with at least one exacerbation.

The relative treatment effect was expressed as an odds ratio (OR) with 95% confidence intervals (95% CI).

The trials were grouped according to the different LABA comparators (indacaterol, salmeterol, and formoterol).

Example 2 (QoL analysis)

Classical pooling for the analyses were SGRQ total score and SGRQ nonresponder rates, as presented by Schmidt et al. (results not shown).

RESULTS

Example 1 (exacerbations analysis)

The IQWiG final report noted an absence of heterogeneity between the groups of placebo-controlled LABA trials.

Clinical endpoints used for the analyses were SGRQ total score and SGRQ nonresponder rates, as presented by Calverley et al. (2007).3

The analysis included 19 randomized, controlled trials (indacaterol, n=2; salmeterol, n=3; formoterol, n=14).

For analysis, the RevMan software was applied. Additionally, a Boehringer Ingelheim internal meta-analysis software was used for recalculation and confirmation of results.

Example 2 (QoL analysis)

For the double-blind (n=12) and open-label (n=3) subgroup analysis, the mean difference was –3.20 (95% CI: –3.75, –2.65).

The IQWiG final report noted an absence of heterogeneity between the groups of placebo-controlled LABA trials.

Heterogeneity: Tau2=0.00; Chi2=2.15, df=9 (P=0.44); I^2=0% (Figure 2).

Subtotal (95% CI) 357 113 3903 100.0 0.94 (0.65, 0.77)

Total (95% CI) 2170 2275 3084 0.71 (0.65, 0.77)

Example 2 (QoL analysis)

The overall effect measure for both SGRQ analyses showed similar behaviour in the open-label and double-blind trials.

The overall effect measures for both SGRQ analyses showed similar behaviour in the open-label and double-blind trials, with the subgroup of double-blind trials expressing a stronger effect in favour of tiotropium.

The interaction Pruska of subgroup and treatment effect was low, indicating that trial design (open label versus double blind) seems to be a significant modifier of the effect of inhaled tiotropium on QoL.

As the modification was quantitative, the effect of inhaled tiotropium on SGRQ total score was substantially underestimated when the administration had been open-label, compared with the gold-standard (double-blind) comparison.

DISCUSSION

The absence of heterogeneity in Example 1 suggests that the benefit of inhaled tiotropium versus salmeterol with regard to preventing COPD exacerbations (demonstrated in direct comparison in the POET-COPD trial) can be extended across all LABAs, as indacaterol, salmeterol and formoterol showed homogeneity in treatment effects when compared to placebo.

In Example 2, the subgroup analysis of the interaction of study design and treatment effect revealed that the magnitude of QoL benefit provided by tiotropium in COPD studies was inconsistent between open-label and blinded administration of tiotropium, with greater benefit seen in blinded trials.

The SGRQ total score analysis was supported and complemented by the nonresponder analysis, which can be interpreted as a sensitivity analysis.

The overall effect measures for both SGRQ analyses showed similar behaviour in the open-label and double-blind trials, with the subgroup of double-blind trials expressing a stronger effect in favour of tiotropium.

The interaction Pruska of subgroup and treatment effect was low, indicating that trial design (open label versus double blind) seems to be a significant modifier of the effect of inhaled tiotropium on QoL.

The modification was quantitative, the effect of inhaled tiotropium on SGRQ total score was substantially underestimated when the administration had been open-label, compared with the gold-standard (double-blind) comparison.

CONCLUSION

These examples demonstrate the useful application of classical meta-analytic methods and a test on subgroup differences to assess heterogeneity across groups of trials based on aggregate trial data.

Most available software packages for meta-analysis allow for easy execution of this test of subgroup differences.

REFERENCES


Acknowledgements

Editorial support was provided by PAREXEL and funded by Boehringer Ingelheim.

Supported by Boehringer Ingelheim.

IPPAR 2015, Milan, Italy.

Figure 1. Meta-analysis of trials comparing tiotropium and placebo (based on SGRQ total score)

Table 1. Results of subgroup analysis (based on SGRQ total score)

Table 2. Results of subgroup analysis (based on SGRQ nonresponder rate)

Table 3. Results of subgroup analysis (based on SGRQ total score)

Table 4. Results of subgroup analysis (based on SGRQ nonresponder rate)