



Meta-Analyse seltener Medikamenten- Nebenwirkungen aus heterogenen klinischen Studien

Ekkehard Glimm, Novartis Pharma

IQWiG im Dialog, 17 Juni 2011

Einführung

- Zunehmende Bedeutung von Meta-Analysen für die Pharmaindustrie
- **Zulassungsbehörden: Medikamentensicherheit**
 - FDA : Nebenwirkungen von NSAIDs, Kardiovaskuläres Risiko von Diabetesbehandlung (2008 FDA Richtlinie)
 - Datentyp Zähldaten: Anteil Patienten mit einer bestimmten Nebenwirkung
- **Kostenträger: Risiko-Nutzen-Bewertung:**
 - IQWiG: Memantin zur Alzheimer-Behandlung
 - NICE
 - Datentyp oft kontinuierlich, z.B. ADCS-Scores

Einführung: Bisherige Erfahrungen

- Kontakte primär mit Zulassungsbehörden (FDA, EMA)
- Primär **Sicherheit**, weniger **Kosten-Nutzen**
- Situation oft: Zusammenfassung relativ weniger, interner Studien
- “individual patient data” (IPD) oft vorhanden
- Wenige Studien, seltene Ereignisse
- FDA: mehrere Analysemethoden, oft “exakte” Methoden als eine Forderung

Methodik

- Statistisches “Standardmodell”:

$$\text{link}(\mu_{ij}) = \alpha_i + \theta_{ij}$$

- μ_{ij} = (wahrer) Effekt von Behandlung j in Studie i
 - α_i = Baseline-Effekt
(z.B. Ereignisrate der Kontrollgruppe in Studie i)
 - θ_{ij} = differentieller Behandlungseffekt in Studie i :
Veränderung von Behandlung j gegenüber Kontrolle
- *link* beschreibt die Skala der Inferenz (e.g. logit für den Odds-Ratio)
 - “Im Prinzip” gleich für Zähl- und kontinuierliche Daten

$$\text{link}(\mu_{ij}) = \alpha_i + \theta_{ij}$$

- Uneingeschränktes Modell ist “saturiert” (=“nutzlos”)
- “Fixed Effects”-Modell: $\theta_{ij} = \theta_i$
 - Baseline-Effekt variiert beliebig
 - Differentieller Behandlungseffekt überall gleich (auf der vorgegebenen Skala!)
- “Random Effects”-Model: $\theta_{ij} \sim N(\theta_i, \sigma^2)$
 - Differentieller Behandlungseffekt variiert von Studie zu Studie
 - Variation folgt vorgegebener Verteilungsannahme

Methodik

- Ziel der Meta-Analyse: Zusammenfassender Schätzer des differentiellen Behandlungseffektes
- Gewichtetes Mittel der Effektschätzungen $\hat{\theta}_{ij}$:

$$\frac{\sum_i w_i \hat{\theta}_{ij}}{\left(\sum_i w_i \right)}$$

- w_i = Gewicht von Studie i

Frage: Welches Gewicht?

Methodik

(kombinierter) Schätzer des Behandlungseffekts

$$\frac{\sum_i w_i \hat{\theta}_{ij}}{\left(\sum_i w_i \right)}$$

Kandidaten für w_i :

- (Studien-spezifische Varianz des Schätzer)⁻¹
- Größe der Studie
- “Shrinkage” in Random-Effects-Modellen:
 $w_i = (z + \text{Studien-spezifische Varianz des Schätzer})^{-1}$,
 $z = \text{Schätzung der Variation zwischen den Studien}$
⇒ Gewichte werden beschränkt

Heterogenität

- Zwei Formen von Heterogenität:
 - **Bei kontinuierlichen und bei Zähldaten:**
Differenzielle Behandlungseffekte θ_{ij} sind nicht gleich (Interaktion von Studie und Behandlung)
 - **Nur bei kontinuierlichen Daten:**
Unterschiedliche *Variation* der Behandlungseffekte in verschiedenen Studien (“Varianzheterogenität”)
- (Hier wird nur die erste Form von Heterogenität besprochen.)

Folgen von Heterogenität

1. Auf den kombinierten Effektschätzer:

- Errechneter Zahlenwert hängt stark von Gewichten ab
- Ignorieren der Heterogenität führt zu engen Konfidenzintervallen

2. Auf die Interpretation:

- Gewichtung hat großen Einfluss
- Starke Abhängigkeit von Methodik untergräbt Glaubwürdigkeit der kombinierten Schätzung

Auffinden von Heterogenität

Q-Test: Summe der Abweichungsquadrate der Einzelschätzungen vom kombinierten Schätzer

- Wenige Studien, seltene Ereignisse: Heterogenität ist sehr schwer aufzufinden
- *Eine sehr alte Kritik:*
Heterogenitätstest versagen oft:
 - Bei wenigen Studien: keine Power
 - Bei vielen Studien: “zu viel Power”: statistische Signifikanz kleiner Unterschiede, die für Interpretation irrelevant
 - Zusätzliches Problem bei Zähldaten: Q-Test ist asymptotisch

Heterogenität bei Zähldaten

- *Differentieller* Behandlungseffekt unterscheidet sich von Studie zu Studie
- Homogenität auf einer Skala bedeutet Heterogenität auf allen anderen:
 - Ist das relative Risiko in allen Studien gleich, so können Risikodifferenz und Odds-Ratio nicht gleich sein.
- Analyse abgeleiteter Behandlungseffekte (z.B. Meta-Analyse von Odds-Ratios als asymptotisch Normalverteilt mit als bekannt angenommenen Varianzen) ist problematisch

Heterogenität bei Zähldaten

Beispiel: Studien des kardiovaskulären Risikos bei Behandlung von Diabetes Mellitus

- Mehrere kleine Studien mit Wirksamkeit als primärer Zielstellung
- Eine große Studie mit myokardialen Ereignissen als primärer Zielstellung

Was ist zu tun, wenn sich hier Heterogenität zeigt?

Heterogenität bei Zähldaten

Seltene Ereignisse: Keine allseits befriedigende Lösung des Heterogenitätsproblems

- Exakte Verfahren (z.B. exakte Version des Mantel-Haenszel-Tests) funktionieren nur im Fixed-Effects-Modell
- Verfahren für Random-Effects-Modelle (Quasi-Likelihood, GEEs) haben generell schlechte asymptotische Eigenschaften
- Bei bayesianischen Verfahren werden die a-priori-Annahmen nicht von den Daten dominiert

Umgang mit Heterogenität

- Random-Effects-Modelle sind kein Allheilmittel:
 - Schwierigkeiten mit der Asymptotik bei Zähldaten
 - Die Annahme über die Random-Effects-Verteilung kann falsch sein
 - Die Modell-induzierte Gewichtung der Einzelschätzer ist u.U. fraglich (gilt aber auch für das Fixed-Effects-Modell !)
- Ein unterschätzter (?) Ausweg:
Aufnahme weiterer Faktoren zur Erklärung der Heterogenität:
 - Phase der Studie
 - Gruppierung der Studien nach den Ein- und Ausschlusskriterien der randomisierten Patienten
 - Bei Vorliegen der individuellen Patientendaten: Modelle mit Studienindex als einer Kovariablen

Einige Thesen

- Heterogenität ist mehr inhaltliches denn statistisch-methodisches Problem
- Heterogenitätsindizes (Q-Test-Statistik, I^2 etc.) sind wichtig, sollten aber informell beurteilt werden und sind nur ein Aspekt von vielen bei der Modellwahl
- Differenzen in den Ansichten zur geeigneten Modellwahl (“fixed vs random effects”) sind meines Erachtens nicht auf formal-mathematischer Ebene lösbar.
- Bei erheblicher Heterogenität muss man die Einzelstudien ansehen. “Random-Effects“-Annahmen sind hierfür kein adäquater Ersatz.

Fazit

- Die statistisch-technischen Mittel zur simultanen Analyse mehrerer Studien stehen zur Verfügung und werden beherrscht.
- Meta-Analysen sind nützlich, wenn die gepoolten Studien bezüglich der Fragestellung “*hinreichend homogen*” sind.
- Bei seltenen Nebenwirkungen ist eine Meta-Analyse oft der einzige Ausweg aus dem Dilemma der geringen Information pro Studie.
- Starke Heterogenität verlangt nach einer inhaltlichen Erklärung.