

Welche wissenschaftlichen Fragen können mit Estimands adressiert werden?



W. Brannath^{1,2}, V. Didelez^{2,3} und M. Scharpenberg^{1,2}

KKSB¹, IfS² und BIPS³



IQWIG im Dialog, Juni 2018

Ziele / Hintergründe des Addendums

- ▶ Intelligenteren und flexibleren Umgang mit Interkurrenten Ereignissen (IE) fördern . . .
... und zwar über das Konzept der “kausalen Estimands”
- ▶ IE soll als Komponente des zu schätzenden Effekt-Parameters (Estimand) betrachtet werden
- ▶ Umgang mit IE in stat. Analyse dient der unverzerrten und effizienten Schätzung der vorab festgelegten Estimands
- ▶ Betrachtung verschiedener Estimands für eine bessere (strukturelle) Sensitivitätsanalyse

Der Umgang mit den zu erwartenden IE wird damit wesentlicher Bestandteil der Studienplanung (weil Bestandteil der Effektparameter)

Strategien im Focus des Addendums

- ▶ *“Treatment-Policy”*:
Endpunkt unabhängig vom IE berücksichtigen (wie bei ITT)
- ▶ *“Composite”*:
IE in den primären Endpunkt integrieren
- ▶ *“Hypothetical”*:
Effekt unter einem hypothetischen Szenario
- ▶ *“Principal Stratum”*:
Effekt in einem IE-spezifischem Patienten-Stratum
- ▶ *“While On Treatment”*:
Effekt bis zum Auftreten des IE (ev. kumulativer Effekt)

Attribute zur Beschreibung eines Estimand

Das Addendum nennt die folgenden Attribute eines Estimand

- A. Population - die anvisierten Patienten
- B. Variablen - Endpunkt(e), die notwendigerweise für jeden Pat. zu erheben sind
- C. Interkurrente Ereignisse (IE) und Umgang mit diesen
- D. Summary auf Populationsebene für den Behandlungsvergleich

Nach Hernán & Scharfstein (2018) fehlt:

A2. Zu vergleichende Behandlungsstrategien/-Pläne/-Policies

Ein *kausaler Estimand* bezieht sich **immer** auf eine *Treatment-Policy*!

Ein grundsätzlicher Widerspruch?

- ▶ In der kausalen Inferenz sind **RCT** für **Treatment-Policies** die Grundlage der Definition eines **kausalen Estimands**.
- ▶ Im Addendum (etc.) bemüht man **kausale Estimands**, um sich vom Dogma der **Treatment-Policy** zu befreien.

Ansprüche an Effektivitätsanalysen

Die Effektivitätsanalysen einer klinischen Studie sollten i.A. (und u.A.)

- ▶ auf implementierbare Behandlungs-Strategien/Policies bezogen sein (Heranán & Scharfstein, 2018),
- ▶ klinisch / medizinisch gut interpretierbar,
- ▶ und hinreichend verallgemeinerbar / in Praxis übertragbar sein.

NRC-Report:

The causal estimands . . . should be meaningful for all study participants, and estimable with minimal assumptions.

In wieweit ist das mit den 5 Strategien des Addendums erreichbar?
Gibt es Alternativen oder ist Zusätzliches notwendig?

Treatment Policy (Endpunkt unabh. vom IE berücksichtigen)

- ▶ Reflektiert das ITT-Prinzip allerdings mit dem (neuen) Anspruch, einen zugehörigen Estimand (mit Treatment Policies) möglichst klar und eindeutig zu definieren
- ▶ Estimand ist dann Effekt von Therapie+IEs (z.B. mit Notfallmedikation, Therapieabbruch und/oder Therapiewechsel)
- ▶ **Potentielles Problem:**
 - Sind das IE und dessen Zusammenhang zu Trt und Outcomes im RCT verallgemeinerbar / auf Praxis übertragbar?
 - Wenn IE zu studienspezifisch, dann Übertragbarkeit schwierig (Frage des Studiendesign)

Treatment Policy (Endpunkt unabh. vom IE berücksichtigen)

▶ *Zu beantworten:*

Was/wie war die Häufigkeit und der Zusammenhang von IE mit Trt und Outcomes?

Wie beeinflussen Veränderungen in Häufigkeit und Zusammenhang von IE mit Trt und Outcomes den Trt-Effekt?

▶ *Schluss:* Benötigen geeignete Sensitivitäts/Robustheitsanalysen

▶ *Unser Vorschlag:* Analysen unter hypothetischen Szenarien mit abweichender Häufigkeit / Zusammenhang (z.B. mittels statistischer Modelle und Inverse Probability Weighting; IPW)

Composite

(IE mit primären Endpunkt kombinieren)

- ▶ Folgt *meist richtiger* Idee, dass IE Teil des Outcomes ist
- ▶ Schwierigkeit liegt in Kombination / Gewichtung zwischen IE und prim. Endpunkt
- ▶ Hierzu existiert Literatur, die *mehr Beachtung* finden könnte
- ▶ Methodische Forschung hierzu sollte intensiviert werden

Beispiele für Methoden:

▶ *Addendum:*

Oder-Verknüpfung von IE mit binärem Endpunkt

▶ *Einfache Methode* (z.B. Gould, 1980; White et al., 2001):

IE in Bezug auf die Werte des Endpunkts Y anordnen
(z.B. IE schlechter als alle Werte des Endpunkts)

→ ordinaler komb. Endpunkt mit pot. Out. O_T und O_K .

Estimand ist z.B. relativer Trt-Effekt: $\Delta := \mathbf{P}(O_T \geq O_K)$

- ▶ *Komplexere Methoden (z.B. Finkelstein & Schoenfeld '99, Buyse '10, Witkowski et al. '04, Rauch et al. '12, Péron et al. '18):*

Endpunkt(e) und IE(s) als multivariaten Outcome: M_T und M_K

Definiere, wann (M_T, M_K) "neutral", "für" oder "gegen T " spricht.

Estimand:

$$\Delta := \mathbf{P}\left((M_T, M_K) \text{ "für } T\right) - \mathbf{P}\left((M_T, M_K) \text{ "gegen } T\right)$$

Composite

(IE mit primären Endpunkt kombinieren)

- ▶ *Vorschlag:* In Sensitivitätsanalysen alternative Kombinationen / Gewichtungen / Ordnungen betrachten
- ▶ *Zur Frage der Verallgemeinerung:* wieder Einfluss von Häufigkeit und Zusammenhang von IE mit Trt und Outcomes mittels hypothetischer Szenarien studieren

Composite – Beispiel zur Illustration

- ▶ Doppel-blinde rand. Studie für einen spez. chronischen Schmerz
- ▶ T vermeintlich verträglicher und wirksamer als K (Standard)
- ▶ Endpunkte = AUC des wöch. Schmerz innerhalb von 3 Monaten
- ▶ Verschiedene Arten von IE:
 - IEa*: Abbruch wegen AE (A-AE), Unwirksamk. (A-U), Tod (M)
 - IEb*: Entscheidung zur Fortsetzung mit K (also T-K oder K-K)
- ▶ *Ordnung*: $M < A-AE = A-U < T-K = K-K < \text{kein IE (Endpkt.)}$
- ▶ *Sens.-Ana.*: $A-AE < A-U$ bzw. $A-AE > A-U$ und $T-K < K-K$
- ▶ *Robustheit*: Rate von *IEb* erhöhen bzw. verringern

- ▶ Abschätzung des Effektes unter einem hypothetischen Szenario (z.B. keine Notfallmedikation verfügbar)
- ▶ Estimand einer implementierbaren Behandlungsstrategie?
- ▶ Wirklich ein sinnvoller Effekt? Hernán & Scharfstein (2018):
*“The E9 Addendum should emphasize the importance of basing regulatory decisions on the comparison of **treatment strategies that may be realistically implemented in well-defined groups of patients.**”*
- ▶ **Stat. Probleme:**
Annahmen / Modelle zur Parameter-Schätzung notwendig
Inwieweit sind die Annahmen willkürlich bzw. überprüfbar?

Unsere Meinungen:

- ▶ Sinnvoll, wenn hyp. Szenario (prinzipiell) umsetzbar und relevant
- ▶ Für Sensitivitätsanalysen gut geeignet
- ▶ Ev. hilfreich, um Spektrum erzielbarer Trt-Effekte abzuschätzen
- ▶ Auch weniger extreme Szenarien betrachten (z.B. mit reduzierter Rate an Notfallmedikationen, Trt-Abbrüchen und -Wechsel)
- ▶ Umgang mit IE muss konsequenterweise zu einem konsistenten Schätzer für angepeilten Estimand führen
- ▶ Imputationsmethode ev. unzuverlässig, wenn sie zusätzliche Annahmen impliziert (z.B. Outcome nach IE in T wie unter K) (Siehe z.B. White et al., 2017).

Inverse Probability of Censoring Weighting

- ▶ Wenn IE von Trt und IE-Historie stochastisch unabhängig:
Analyse der IE-freien Patienten (PP-Analyse) liefert unverzerrten Schätzer für hypothetischen Effekt ohne IE.
- ▶ Können mittels IPCW (unter Modellannahmen) Studiendaten auf das Unabhängigkeitsszenario umgewichten
(siehe z.B. Robins et al., 200; Robins & Finkelstein, 2000)
- ▶ Die Gewichte werden auf Basis aller Patienten bestimmt.
- ▶ Befreit uns von Imputationsmaßnahmen und -annahmen.
- ▶ Das Unabhängigkeitsszenario kann näher am realen Szenario liegen, als das durch eine spezifische Imputation erzeugte.

IWCP – illustratives Beispiel

- ▶ Dieselbe Studie für spez. chr. Schmerz nun mit zusätzlichem IE:
IEc: Studienabbruch aus anderen / unbekanntem Gründen
- ▶ *Estimand*: Behandlungseffekt ohne *IEc*
- ▶ Logistische Regression für $\mathbf{P}(\text{kein } IEc \mid \text{Historie und Trt})$
Historie = z.B. vorangegangener Schmerz und AEs
- ▶ ANOVA mit Patienten ohne *IEc* und Gewichten (skizzenhaft)

$$w = \frac{\mathbf{P}(\text{kein } IEc) \mathbf{P}(\text{Historie und Trt})}{\mathbf{P}(\text{kein } IEc \mid \text{Historie und Trt}) \mathbf{P}(\text{Historie und Trt})}$$

(*Marginal Structural Models*, Robins et al. 2000)

- ▶ *Bemerkung*: Multiple Imputation benötigt Modell für bedingte Verteilung des Endpunkts

IWCP – illustratives Beispiel

- ▶ Dieselbe Studie für spez. chr. Schmerz nun mit zusätzlichem IE:
IEc: Studienabbruch aus anderen / unbekanntem Gründen
- ▶ *Estimand*: Behandlungseffekt ohne *IEc*
- ▶ Logistische Regression für $\mathbf{P}(\text{kein } IEc \mid \text{Historie und Trt})$
Historie = z.B. vorangegangener Schmerz und AEs
- ▶ ANOVA mit Patienten ohne *IEc* und Gewichten (skizzenhaft)

$$w = \frac{\mathbf{P}(\text{kein } IEc)}{\mathbf{P}(\text{kein } IEc \mid \text{Historie und Trt})}$$

(*Marginal Structural Models*, Robins et al. 2000)

- ▶ *Bemerkung*: Multiple Imputation benötigt Modell für bedingte Verteilung des Endpunkts

IWCP – illustratives Beispiel

- ▶ Dieselbe Studie für spez. chr. Schmerz nun mit zusätzlichem IE:
IEc: Studienabbruch aus anderen / unbekanntem Gründen
- ▶ *Estimand*: Behandlungseffekt ohne *IEc*
- ▶ Logistische Regression für $\mathbf{P}(\text{kein } IEc \mid \text{Historie und Trt})$
Historie = z.B. vorangegangener Schmerz und AEs
- ▶ ANOVA mit Patienten ohne *IEc* und Gewichten (skizzenhaft)

$$w = \frac{\mathbf{P}(\text{kein } IEc)}{\mathbf{P}(\text{kein } IEc \mid \text{Historie und Trt})}$$

(*Marginal Structural Models*, Robins et al. 2000)

- ▶ *Bemerkung*: Multiple Imputation benötigt Modell für bedingte Verteilung des Endpunkts

Principal Stratum (Effekt in IE-spezifischem Pat.-Stratum)

- ▶ Effektivität in einer **Sub-Population (Stratum)**, die durch (die) Baseline-Variablen **nicht identifizierbar** ist
Z.B.: alle Pats, die unter T und K keine Notfallmedikation benötigen werden; oder: die weder T noch K abbrechen werden
- ▶ Estimand ist noch virtueller als mit *Hypothetical* Strategie
- ▶ Basiert auf keiner implementierbaren Behandlungsstrategie
- ▶ Daher in Fachkreisen der kausalen Inferenz *umstritten*; (z.B. Pearl 2011; David & Didelez 2012; weitere in *Int J Biostat* 2011/12)
- ▶ **Statistische Schwierigkeit:**
Schätzung basiert auf unüberprüfbare Annahmen; z.B. kein Pat. benötigt Notfallmedikation unter K aber nicht unter T .

Principal Stratum (Effekt in IE-spezifischem Pat.-Stratum)

Unsere Meinung:

- ▶ Ggf. zur Untersuchung wiss. / strategischer Fragen; Z.B.:
 - Ist die neue Therapie in einer gewissermaßen “optimalen” Subpopulation überhaupt effizient (genug)?
 - Wirkt eine Therapie auch bei Patienten, die keinen spezifischen transienten Zustand entwickeln (Baldur Magnusson, Novartis)
- ▶ Ungeeignet für regulative und HTA-Fragen (Hér. & Sch., 2018)
- ▶ Estimand liefert i.A. keinen reinen “de jure” Effekt:
Behandlungs- und Beobachtungsunterschiede, die sich zwischen Random. und IE ergeben, bleiben unberücksichtigt (Bei offenen Studien eventuell sogar verzerrter Vergleich)

While on treatment

(Effekt bis zum Auftreten des IE)

- ▶ Ist letztlich Frage des (prim.) Endpunktes und sollte so diskutiert werden:

Genügt es aus *medizinischer (und Patienten-) Sicht*, den Trt-Erfolg nur bis zum IE zu bewerten?

- ▶ In spezifischen Fällen ausreichend, oft / meist aber wohl nicht
- ▶ Auch kein reiner “de jure” Effekt (bei offenen Studien verzerrt)
- ▶ **Weitere Probleme:**
 - Beurteilen eventuell nur kurzfristige Effekte
 - Ungleichmäßiges Follow-up
 - Gewichtung von Endpunkt und Zeit bis IE ev. widersprüchlich (Severity-Endpunkt → frühes IE führt zu besserem Outcome)

Diskussion

- ▶ Zu **jedem** kausalen Estimand gehören wohl-definierte Treatment-Policies, ...
... die für regulatorische und HTA Fragen praktisch umsetzbar und medizinisch relevant sein sollen
- ▶ Principal Strata sind **nicht** identifizierbar und damit für regulatorische und HTA-Fragen ungeeignet
- ▶ IE kann Teil der Treatment-Policies und/oder Endpunkte sein
Für beides stellt sich die Frage nach med. Relevanz und Verallgemeinerbarkeit (Sensitivitäts / Robustheitsanalysen)
- ▶ Hypothetisches Szenarien sinnvoll, wenn sie zu medizinisch relevanten und umsetzbaren Behandlungsstrategien gehören
- ▶ *Wichtiger Aspekt:* Konsequenzen für das Studiendesign (ggN)

Ausgewählte Literatur – Vielen Dank!

- ▶ Hernán MA, Scharfstein D (2018) Cautions as Regulators Move to End Exclusive Reliance on Intention to Treat. *Ann Intern Med* 168:515-6
- ▶ Péron J, Buyse M, Ozenne B, Roche L, Roy P (2018) An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. *Stat Methods Med Res* 27: 1230-9
- ▶ Geraldine Rauch G, Jahn-Eimermacher A, Brannath W, Kieser K (2014) Opportunities and challenges of combined effect measures based on prioritized outcomes. *Stat Med* 33:1104-20
- ▶ Fischer K, Goetghebeur E, Vrijens B, White IR. (2011) A structural mean model to allow for noncompliance in a randomized trial comparing 2 active treatments. *Biostatistics* 12:247-57
- ▶ Dawid P, Didelez V (2012) “Imagine a Can Opener” – The Magic of Principal Stratum Analysis. *Int J Biostat* 8(1) Art. 19
- ▶ Pearl J (2011) Principal Stratification – a Goal or a Tool? *Int J Biostat* 7(1) Art. 20
- ▶ Robins JM, Finkelstein DM (2000) Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 56:779-88.
- ▶ Robins JM, Hernán MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550-60.