



Wien, 31.10.2016

Betreff: Kommentierung des Endberichts ‚Ambulante kardiologische Rehabilitation Teil VI: Multizentrische, prospektive Kohortenstudie mit zwei Vergleichsgruppen (mit/ohne Phase III)‘ von Prof. Dr. Pölz und Dr. Schimetta vom 26.9.2016

Als AutorInnen des Berichts erlauben wir uns im Folgenden zu den im o.g. Schreiben geäußerten Kritikpunkten Stellung zu nehmen:

Ad 1.) Hauptzielparameter-Auswertung:

Eine reine Zufallsrandomisierung stellt sicher, dass sich Interventions- (IG) und Kontrollgruppe (KG) in allen Basisvoraussetzungen nur zufällig unterscheiden. Damit lässt sich der Effekt einer Intervention biasfrei, d.h. ohne **systematische** Verzerrung schätzen, und im Fall einer "intention to treat"-Analyse kausal interpretieren¹.

In einer nicht (voll) randomisierten Studie besteht die Gefahr, dass PatientInnen mit bestimmten Ausgangsvoraussetzungen tendenziell eher in die IG bzw. KG kommen. Sofern solche Ausgangsvoraussetzungen potentiellen Einfluss auf das Ergebnis haben, verzerrt diese Zuordnung die Schätzung des Effekts **systematisch**².

Eine potentielle systematische Verzerrung der Effektschätzung kann für nicht gemessene, und von erfassten Merkmalen unabhängige Parameter durch keine Analysemethode ausgeschlossen werden.

Zur Kompensation eines potentiell verzerrenden Einflusses **erhobener** Merkmale bietet der Propensityscore ([Rosenbaum & Rubin, 1983](#) sowie [Rosenbaum & Rubin, 1984](#)) einen methodischen Zugang.

Der Propensityscore schätzt die Wahrscheinlichkeit einer Patientin/ eines Patienten mit bestimmten Ausgangswerten, der IG zugeordnet zu werden. In weiterer Folge nimmt man an, dass innerhalb von PatientInnen mit identischer Zuordnungswahrscheinlichkeit de facto eine zufällige Zuordnung zu IG bzw. KG erfolgt³.

Zur Bildung eines Propensityscores bietet sich eine logistische Regression an. Der resultierende Propensityscore wird dabei von verschiedenen Aspekten der Modellbildung⁴ beeinflusst. Um die Transparenz der Auswertung sicherzustellen und etwaige Beliebigkeiten bei der Auswertung auszuschließen, wurde in diesem Projekt die Vorgangsweise im Studienprotokoll (siehe <http://eprints.hta.lbg.ac.at/971/>) genau definiert und dementsprechend vorgegangen.

Im gegenständlichen Fall hatte das zur Folge, dass für den Effektparameter einer Kategorie des Erwerbsstatus ein extremer Wert geschätzt wurde. Pölz/Schimetta⁵ haben recht, wenn sie feststellen,

-
- 1 Die grundsätzlichen Problematiken des induktiven Schlusses sowie der inferentiellen Statistik werden davon nicht berührt.
 - 2 In einer nicht manipulationssicher randomisierten Medikamentenstudie wurde ein vielversprechendes neues Antibiotikum mit einem Standardprodukt verglichen. Die behandelnden Ärzte hatten die Neigung unter Umgehung der Randomisierung, Patienten in besonders kritischem Zustand das neue Produkt zukommen zu lassen. Das hatte einen massiv verzerrenden Einfluss auf den geschätzten Effekt, sodass die Studie abgebrochen wurde.
 - 3 Diese Annahme ist nicht prüfbar, da sie mit der Voraussetzung verbunden ist, dass alle effektrelevanten Ausgangswerte erfasst wurden.
 - 4 Unter anderem haben die Kodierung und die Transformation von Merkmalen, sowie die Aufnahme von Wechselwirkungsparametern unterschiedlicher Ordnung Einfluss auf den Score.
 - 5 W. Pölz/ W. Schimetta, Kommentierung des Abschlussberichts der Phase III Rehabilitations-Studie

dass dieser extreme Wert die anderen Parameterschätzungen beeinflussen kann. Es wäre demnach kritisch, den Einfluss einzelner Parameter zu interpretieren⁶.

Es war vielleicht ein Fehler unsererseits, nicht darauf hinzuweisen, dass der extreme Wert durch die formale Parametrisierung des Modells zustande kommt und zB durch andere Bildung der Kontraste oder Orthogonalisierung vermeidbar ist. Der spezifische Einfluss einzelner Merkmale auf die Zuordnung ist, isoliert betrachtet, für die Hauptfragestellung der Studie jedoch nicht von Bedeutung und auf den Propensityscore selbst hat die Parametrisierung keinen Einfluss⁷.

Wie oben dargestellt, schätzt der Propensityscore die Zuordnungswahrscheinlichkeit der PatientInnen. Der gebildete Propensityscore beschreibt diese Wahrscheinlichkeit in gutem Maß. Der relative Anteil an IG-PatientInnen in den fünf Propensitygruppen steigt von 24% über 33%, 34%, 75% auf 93% in der höchsten Gruppe. Wir stimmen daher der Ansicht, dass die Ergebnisse "sinnlos" bzw. "falsch" wären im Hinblick auf die Anwendung des Propensityscores nicht zu.

Ebenso erübrigt sich eine Neuberechnung des Propensityscores "unter Ausschaltung des aufgetretenen Fehlers", da der vermeintliche "Fehler" tatsächlich keinen Einfluss auf den Propensityscore hat⁸. Daraus folgt, dass sich der angegebene Kohortenunterschied durch eine Neuberechnung nicht ändern würde. Aus diesem Grund ist Multikollinearität für die Anwendbarkeit eines Propensityscores nicht von Bedeutung ([Krug, 2014](#), [Fang et al., 2010](#), [Gripencrantz, 2014](#))⁹.

Unabhängig davon, dass das Studienprotokoll keine Alternative zuließ, waren wir aber an der Robustheit der Ergebnisse interessiert und haben verschiedene, nicht vorgesehene Auswertungen vorgenommen, die unser Vertrauen in die Schlussfolgerung des Berichts bestärken. A priori hatten wir keine Erwartung bezüglich einer Richtung eines Bias. Für uns waren neben einer nicht verzerrten Zuordnung sowohl eine für die IG günstige, als auch eine für die KG günstige Tendenz vorstellbar. Wir haben daher explorativ auch Prognosescores gebildet, wobei diese zeigten, dass PatientInnen mit günstiger Prognose eher der IG zugeordnet wurden. Ebenso haben wir die Ergebnisse des- nach Propensityscore stratifizierten- Tests für den primären Endpunkt mit multiplen Regressionen unter Einschluss der Ausgangsmerkmale verglichen.

Als Nachtrag mit Rücksicht auf den Kommentar von Pölz/Schimetta wurde eine Version des Propensityscores ohne das Merkmal Erwerbsstatus gebildet und der primäre Endpunkt danach stratifiziert ausgewertet. Die Koeffizienten der anderen Merkmale ändern sich großteils nur geringfügig. Die Korrelation zwischen den betas des Original-Propensityscores und jenen des Propensityscores ohne das Merkmal Erwerbsstatus beträgt $r = 0,88$. Die Zuordnung zu den fünf Propensitykategorien ändert sich geringfügig. Die Korrelation zwischen den Original-Propensitykategorien und jenen ohne Berücksichtigung des Erwerbsstatus betragen $r = 0,93$ und $\tau = 0,88$. Für den primären Endpunkt entsprechen die Ergebnisse im Wesentlichen den Originalergebnissen. Trotz Nichtberücksichtigung des Erwerbsstatus beträgt der adjustierte Mittelwertsunterschied $-0,08$ und der p-Wert des stratifizierten Test $p=0,560$. Bei den hier angeführten Ergebnissen ist daran zu erinnern, dass sie nicht protokollkonform zustande gekommen sind, da der

1. Version v. 26.9.2016

- 6 Extreme Parameterschätzung, die durch Multikollinearität zustande kommen, beeinflussen zwar die Varianz anderer Parameter, ein Bias lässt sich davon aber nicht ableiten [Weitzen et al., 2004](#), [Harrell, 2001](#), [Harding, 2003](#), [Rubin, 2004](#). Grundsätzlich wäre die Interpretation einzelner Parameter bei ausgeprägter Multikollinearität immer schwierig [Harrell, 2001](#).
- 7 Erwerbsstatus ist ein nominales Merkmal, das je nach Wahl der Kontraste zu unterschiedlichen Parameterschätzungen führt. Auch durch eine Orthogonalisierung der Prädiktoren über eine Hauptkomponentenanalyse lässt sich eine Multikollinearität vermeiden, was zu weniger extremen Parameterschätzungen führt.
- 8 Der Propensityscore errechnet sich als gewichtete Summe der berücksichtigten Merkmale, wobei die Merkmale mit den geschätzten Effektparametern gewichtet wurden. Formal unterschiedliche Parametrisierungen oder Orthogonalisierung führt zu unterschiedlichen Effektparametern, ohne dass sich die gewichtete Summe, also der Propensityscore insgesamt dadurch ändert.
- 9 Kollinearität beeinflusst nicht den gemeinsamen Effekt von hochkorrelierten Merkmalen oder eine Prognose aufgrund hochkorrelierter Merkmale [Harrell, 2001](#).

Ausschluss des Erwerbsstatus inhaltlich nicht gerechtfertigt ist¹⁰. Eine detailliertere Darstellung der Ergebnisse ziehen wir im Rahmen des noch unfertigen Tabellenbands in Betracht. Im Hauptbericht halten wir die gegenwärtige Darstellung für ausreichend, da, wie oben erwähnt, eine Interpretation der Einzelparameter nicht in Zusammenhang mit der Hauptfragestellung steht.

Ad 2.) Hauptzielparameter-Auswahl:

Es mag sein, dass die Erläuterungen zur Berechnung des primären Endpunktes im Bericht nicht ausreichend leicht lesbar sind. Insbesondere scheint die Unterscheidung zwischen den sechs Rehabilitationszielkategorien und den 10+2 Einzelrehazielen nicht für jede Leserin/ jeden Leser deutlich genug gemacht worden zu sein. Die Kommentierungen (Pözl/ Schimetta⁵) dürften darauf beruhende Missverständnisse bezüglich der Definition des primären Endpunktes enthalten.

Wie im Bericht¹¹ in Punkt 2.4.1 und Tabellen 2.4-1 sowie 2.4-2 dargestellt, werden sechs, der neun in den Guidelines (Benzer, 2008)¹² angeführten Risikokategorien sowohl im Rahmen der Einschlusskriterien, als auch bei der Definition des primären Endpunktes berücksichtigt. Aus diesen sechs Risikokategorien leiten sich 10 Einzelziele ab, die um zwei Veränderungskategorien (für Leistungsfähigkeit und BMI) erweitert wurden.

Als **Einschlusskriterium** wurden mindestens drei offene Risikokategorien festgelegt, um bei jeder Patientin/ jedem Patienten ein breites Spektrum für positive Veränderungen sicherzustellen. Der **primäre Endpunkt** hingegen beruht im Interesse einer höheren Trennschärfe auf insgesamt zwölf Einzelzielen. Demzufolge ist die Feststellung *"Es handelt sich um eine diskrete Variable mit 13 Zahlenwerten (von -6 bis +6), ..."* nicht zutreffend. Der primäre Endpunkt hat unter Berücksichtigung der Einschlusskriterien einen theoretischen Wertebereich von -10 bis 9, umfasst also 20 Werte. Tatsächlich aufgetreten sind Werte zwischen -5 und +5.

"... die meisten PatientInnen (>90%) haben aber nur einen Spielraum von 3 bis 4 Verbesserungen ..." dürfte sich auf die Tabelle 3.1-16 auf Seite 54 des Berichts beziehen, in der die Zahl offener Risikokategorien dargestellt wird. Tatsächlich haben knapp 93% der PatientInnen bei Studieneinschluss nur drei oder vier offene Risikokategorien. Der Anteil von PatientInnen mit nur drei offenen Einzelzielen beträgt jedoch 56%. Der Mittelwert offener Einzelziele beträgt 4.8, wodurch der primäre Endpunkt eine relativ gute Auflösung in beiden Richtungen bietet. Angemerkt sei, dass in der Projektvorbereitung seitens der Leistungserbringer unterstrichen wurde, dass man sich von der Phase III vor allem eine Stabilisierung des Endzustandes der Phase II erwarte, sodass aus Sicht der Leistungserbringer vor allem die Differenzierung im Verschlechterungsbereich bedeutsam war.

Wie im Bericht dargestellt, wurden die Risikokategorien den nationalen Guidelines entnommen, die auch auf der Seite der AGAKAR http://agakar.at/guidelines/oekg_guidelines.pdf (Zugriff am 21.10.2016) zu finden sind. Die Vermutung „... , wobei 3 der 6 Zielkategorien - "Blutdruck", "Blutfette" und "Blutzucker" - medikamentös beeinflussbar bzw. optimierbar sind (da geht es zumeist eher um den Medikationsaufwand bis zur Zielerreichung als um die Zielerreichung per se) und auch die Zielkategorien "Rauchen" und "Körpergewicht" (mutmaßlich, ohne den Kardiologen vorgereifen zu wollen) nur indirekt mit einer ambulanten Rehabilitation zu tun haben, bleibt also nur "Leistungsfähigkeit" als direkt beeinflussbare Zielkategorie.“ dürfte nicht dem Konzept der Phase III Rehabilitation entsprechen, auch weil, den Guidelines entsprechend (Benzer, 2008), die

10 Es sei auch darauf hingewiesen, dass eine Nachauswertung, die nach Kenntnis des Ergebnisses gefordert wird, das Risiko eines Bias in sich trägt, da die Motivation mit einem unerwünschten Ergebnis zusammenhängen kann. Dabei ist es unerheblich, ob eine betroffene Partei selbst oder jemand von ihr beauftragter die Forderung erhebt. Eine sachliche Rechtfertigung würde daran nur dann etwas ändern, wenn sichergestellt wäre, dass diese in jedem Fall, also auch bei einem für die Partei erwünschtem Ergebnis erhoben worden wäre. Forderungen, die auf Nichtbefolgung des a priori bekannten Studienprotokolls abzielen, sind besonders fragwürdig.

11 Endbericht Ambulante kardiologische Rehabilitation Teil VI (verfügbar unter <http://eprints.hta.lbg.ac.at/1101/>)

12 siehe http://agakar.at/guidelines/oekg_guidelines.pdf

medikamentöse Sekundärprävention¹³ sowie die Entwöhnung vom Rauchen zu den Zielwerten gehören.

Soweit wir über den Rahmenvertrag zur Phase III Rehabilitation informiert sind, sieht er rund 70% des Gesamtaufwandes einer Phase III für Trainings- und Bewegungstherapie vor. Andere therapeutische Interventionen betreffen Psychokardiologie, Ernährungsschulung, Entspannungstraining, theoretische Schulung und Raucherentwöhnung (siehe beispielsweise <http://www.gesundheit-wels.at/Cardio-Vital-Wels/Ambulante-kardiologische-Rehabilitation.html?hp=11>).

Insofern ist uns die Feststellung "*Dieses Manko betrifft vor allem PatientInnen aus der Interventionskohorte, die da auf Verbesserungen abzielt.*" nicht verständlich.

Am Rande erwähnt sei, dass wir auch einen alternativen Endpunkt gebildet und ausgewertet haben. Dieser wurde, abgesehen von der Berücksichtigung von Harnsäure¹⁴ entsprechend eines Vorschlags der AGAKAR "sekundäre Studienendpunkte_EK_12 03 2012a.doc" vom 8.6.2012 (Email Prim. Harpf) berechnet. Dieser alternative Endpunkt korreliert mit dem primären Endpunkt mittelhoch ($r=0,69$, $\tau=0,56$, $\rho=0,69$). Der nicht adjustierte Unterschied beträgt dabei $-0,14$, bei einem nicht stratifizierten p-Wert von $0,937$ (im Gegensatz zu $-0,29$ für den primären Endpunkt), der adjustierte Unterschied fällt zuungunsten der IG ($0,14$, $p=0,533$) aus. Da diese Auswertung im Studienprotokoll nicht vorgesehen war, und der Endpunkt durch das Fehlen der Harnsäuredaten nicht genau dem Vorschlag entsprach, hielten wir eine Erwähnung der Ergebnisse nicht für angebracht.

Ad 3.) Intent-To-Treat (ITT)-Analyse:

Im Studienprotokoll¹⁵ wurden die Regeln der Auswertung definiert. Dort wird im Punkt 2.10.1 dargestellt, dass eine maximale Drop-out-Rate von 33% erwartet wird, und deswegen um 50% mehr PatientInnen in die Studie aufgenommen werden sollen, als am Ende ausgewertet würden.

Weiters wurde festgelegt, dass es für die definitive Zuordnung zur Interventionsgruppe erforderlich sei, die Phase III zumindest anzutreten (siehe 2.12). Auch im Bericht (Punkt 2.6.4 und Tabelle 3.1-3) wurde das dargelegt. Dahinter stand die Überlegung, dass der Einfluss der Phase III durch PatientInnen, die zwar einwilligten, aber dann, zB wegen fehlender Bewilligung durch den SV-Träger, nie in Kontakt mit einer Phase III Rehabilitation gekommen sind, unbillig verschleiert werden könnte. Da auch nicht auszuschließen war, dass initiale KG-PatientInnen ohne Rücksicht auf die Studienzuordnung eine Phase III Rehabilitation absolvieren, und so, bei Wirksamkeit der Phase III, das Ergebnis der Kontrollgruppe unangebracht jenem der Interventionsgruppe annähern würden, wurden solche PatientInnen für die Analyse der Interventionsgruppe zugeordnet.

Es mag als irreführend empfunden werden, dass einmal im Text der Marginalspalte "ITT-Analyse" zu lesen ist. Im Text wird jedoch im Sinne des CONSORT-Statements, 2010 ([Schulz et al., 2010](#), checklist, Punkt 16) erläutert, wie die finale Gruppenzuordnung vorgenommen werden sollte und wurde. Es erfolgte somit keine post hoc Änderung des Auswertungsplans im Sinne einer Abkehr von einer "echten" ITT-Analyse, da eine Auswertung von PatientInnen mit fehlendem Follow-Up nur bei Verstorbenen vorgesehen war (Punkt 2.12).

Ein tatsächlicher Mangel des Studienprotokolls bestand darin, dass zwar festgelegt wurde, dass fehlende Einzelwerte der Abschlussuntersuchung ersetzt würden, dass an der Stelle jedoch kein Maximalumfang fehlender Werte definiert worden war. Dieser Maximalwerte wurde dann vor Berücksichtigung der Gruppenzuordnung mit der Hälfte der Einzelkategorien festgelegt. Ohne eine Obergrenze beim Ersatz fehlender Einzelwerte wären alle PatientInnen mit Ausgangswerten auswertbar gewesen. Damit wären mehr als 50% der berechneten Veränderungen ohne Daten der

13 Der Medikamentenkonsum wurde in der Studie zwar erhoben und untersucht, aber nicht im primären Endpunkt berücksichtigt, da in der Projektvorbereitung davon ausgegangen worden war, dass praktisch alle Patienten am Ende der Phase II korrekt mit Medikamenten versorgt wären, und somit nur negative Veränderungen damit erfassbar wären.

14 Harnsäure wurde nicht erhoben

15 Studienprotokoll Ambulante kardiologische Rehabilitation Teil V, siehe auch [Piso et al., 2012](#)

Enduntersuchung zustande gekommen, was aus unserer Sicht eine Interpretation im Sinn der Fragestellung unmöglich gemacht hätte. Abgesehen von diesem Extrem war die Festlegung eher unkritisch, denn von einer niedrigeren Schranke wäre nur eine geringe Zahl von PatientInnen betroffen gewesen¹⁶.

Die Feststellung "... eigentlich hätte ITT eine Fallzahl (n) von 157 bzw. 199 haben müssen, so wurde es ein n von 76 bzw. 81." wäre unter der Voraussetzung einer bestimmten Definition von ITT¹⁷ korrekt, sie trifft auf die protokollgemäße Auswertung der Studie jedoch nicht zu.

Der Feststellung "Da die (meisten) Missing Values durch den Median der Realwerte der Kohorte ersetzt wurden, hätte ein komplettes bzw. korrektes ITT-n zwar vermutlich keinen größeren Kohortenunterschied erbracht, allerdings wahrscheinlich die Varianz massiv verringert." stimmen wir bezüglich der Konsequenzen für Kohortenunterschied und Varianz formal zu.

Es ist zu bedenken, welchen Zweck die Bestimmung der Stichprobenvarianz verfolgt. Im Zusammenhang mit der Bildung von Konfidenzintervallen oder der Prüfgröße eines Signifikanztests dient die Stichprobenvarianz als Schätzwert für die unbekannte Populationsvarianz. Eine massive Verringerung der standardmäßig berechneten Stichprobenvarianz durch eine Medianimputation wäre auf eine, unter solchen Bedingungen für inferentielle Zwecke ungeeignete Varianzberechnung zurückzuführen¹⁸.

Ad 4.) Per-Protocol (PP)-Analyse:

Vor der Diskussion einer per-protocol-Analyse sei darauf hingewiesen, dass sich die Feststellung, "So ist das Minimum an abgerechneten Reha-Trainingseinheiten 1 (!), ..." vermutlich auf die Angaben in Tabelle 3.1-4 des Berichtes bezieht. Wie dort im Text vermerkt, beziehen sich diese Werte auf jene 147 Fälle, von denen Kostenträger-Angaben zur Phase III vorliegen. Bei den, für den primären Endpunkt Auswertbaren betrug das Minimum hingegen 11,5 Reha-Einheiten. Achtzig Prozent haben mindestens 55 Einheiten konsumiert und der Median lag bei 67,5 Einheiten¹⁹.

Es stimmt, dass, wie in der Kommentierung angemerkt, kürzere Teilnahme geringere Kosten verursacht, was in der ökonomischen Analyse im Projektteil VI auch berücksichtigt wird.

Unsererseits wurde eine per-protocol-Analyse für den Bericht nicht in Betracht gezogen, da sie aus unserer Sicht kaum im Sinn der Hauptfragestellung interpretierbar wäre. Eine protokollgemäße Intervention dauert rund sechs bis neun Monate. Beim gegebenen PatientInnenkollektiv ist es dabei nicht unwahrscheinlich, dass es zu einem Ausscheiden wegen einer Verschlechterung oder eines Reevents kommt. Auch war nie geplant, Daten zu erheben, die bei einer per-protocol-Analyse zur Vermeidung eines guarantee-time-bias²⁰ erforderlich gewesen wären. Ein Ergebnis zugunsten der Intervention hätte dadurch geringe Evidenz. Ein für die Phase III ungünstiges Ergebnis würde weniger leicht durch selektives Ausscheiden der besonders "guten" PatientInnen erklärbar sein. In Summe erschiene uns eine per-protocol-Analyse daher zur fairen Beurteilung der Phase III ungeeignet.

16 Welche Auswirkung eine geänderte Obergrenze auf das Hauptresultat gehabt hätte, konnten wir zum Zeitpunkt der Festlegung nicht feststellen, da damals die Daten der finalen Zuordnung noch nicht vorlagen. Später wollten wir es im Interesse unserer Unvoreingenommenheit nicht untersuchen.

17 Zur Definition von Intention-To-Treat gibt es, besonders im Zusammenhang mit fehlenden Werten, unterschiedliche Definitionen [Alshurafa et al., 2012](#), [Bell et al., 2014](#), [Gupta, 2011](#), [Hollis & Campbell, 1999](#), [Johnston & Guyatt, 2016](#), [Little & Kang, 2015](#). Altman et al. 2001 [Altman et al., 2001](#) stellen noch fest: "Although those participants cannot be included in the analysis, it is customary still to refer to an intention-to-treat analysis.". 2010 schreiben sie im CONSORT statement [Schulz et al., 2010](#): "We replaced mention of 'intention to treat' analysis, a widely misused term, ..."

18 siehe zB [Barzi & Woodward, 2004](#) oder [Zhang, 2016](#)

19 Diese Angaben sind nicht im Bericht enthalten, sondern für den Tabellenband vorgesehen.

20 siehe zB [Lash & Cole, 2009](#), [Hanley & Foster, 2014](#)

Ad 5.) Drop Outs:

Die Verwunderung über die hohe Drop-Out-Rate teilen wir. Wie im Bericht unter Punkt 3.1.3 und in Tabelle 3.1-5 dargestellt, wurde versucht, dem entgegenzuwirken.

Wie im Bericht erwähnt, meinen auch wir, dass ein Attrition Bias nicht auszuschließen ist. Ebenso stimmen wir mit Pölz/Schimetta in der Feststellung "... dann ist ein deutliches Abweichen des Studienergebnisses von der Wahrheit keinesfalls auszuschließen und vielmehr konkret in Erwägung zu ziehen" in gewissem Sinn überein²¹. Ein Hinweis, in welche Richtung die mögliche Abweichung gehen könnte, lässt sich aus unserer Sicht aus den Daten jedoch nicht ableiten. Auch ist uns kein plausibles Argument bekannt, das auf eine bestimmte Richtung schließen ließe.

Für unzutreffend halten wir aber die Feststellung "... dass von mehr als der Hälfte der Studienteilnehmer die von Anfang an bekannten Follow-Up-Verpflichtungen negiert worden sind ...". Im Protokoll wird unter 2.10.2 angeführt, dass die Studienteilnahme jederzeit beendet werden kann, was auch die Deklaration von Helsinki ([WMA, 2013](#))²² fordert. Demzufolge ist es irreführend, von einer Follow-Up-Verpflichtung zu sprechen.

Ad 6.) Design:

"Da der Hauptverband Österreichischer Sozialversicherungsträger als Auftraggeber fungierte und es offensichtlich dessen Ziel war, die Sinnhaftigkeit einer bisher finanzierten Therapie (ambulante Phase III Cardio-Reha) zu hinterfragen, ist zu diskutieren, ob bei dieser Konstellation nicht ein Nichtunterlegenheitsansatz (Null-therapie nicht schlechter als ambulante Phase III Cardio-Reha) zu wählen gewesen wäre."

Nach unserem Informationsstand gab es zum Zeitpunkt der Studienplanung eine temporäre Vereinbarung zur Kostenübernahme zwischen dem Hauptverband und den Anbietern der Phase III Rehabilitation. Die Vereinbarung sah eine Untersuchung des Effekts der Phase III vor einer (unbefristeten) Verlängerung vor.

Die Vereinbarung war temporär begrenzt, da die Intervention nicht a priori als wirksam betrachtet worden war. Aus der Tatsache einer vorläufigen Finanzierung lässt sich aus unserer Sicht keine Umkehr der erforderlichen Evidenz ableiten.

Im Studienprotokoll wird unter "Rechtfertigung/Relevanz der geplanten Studie" festgehalten, dass es keine ausreichende Evidenz für die Wirksamkeit der Phase III gebe. Folglich erschien eine Superiority-Studie erforderlich.

"Hätte der Hauptverband Österreichischer Sozialversicherungsträger die betroffenen Kardiologen (AGAKAR) aufgefordert, Daten zum Wirksamkeitsnachweis zu liefern, hätte die AGAKAR als Auftraggeber eine entsprechende Studie durchführen müssen, ..."

Soweit wir informiert sind, war vereinbart, dass der Hauptverband eine entsprechende Studie durchführt. Abgesehen davon erschließt sich uns nicht, inwiefern eine von der AGAKAR in Auftrag gegebene Studie grundsätzlich eher einer korrekten und fairen Beurteilung des Effekts der Phase III Rehabilitation dienen sollte, als eine mit dem Hauptverband als Auftraggeber.

Brigitte Piso und Heinz Tüchler

21 Dabei sehen wir von der Problematik des Wahrheitsbegriffes ab.

22 siehe Punkt 26 der Deklaration

Literatur:

- [Alshurafa et al., 2012] Alshurafa, M., Briel, M., Akl, E. A., Haines, T., Moayyedi, P., Gentles, S. J., Rios, L., Tran, C., Bhatnagar, N., Lamontagne, F., Walter, S. D., and Guyatt, G. H. (2012). Inconsistent Definitions for Intention-To-Treat in Relation to Missing Outcome Data: Systematic Review of the Methods Literature. *PLoS ONE*, 7(11):e49163.
- [Altman et al., 2001] Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., GÄtzsche, P. C., Lang, T., and CONSORT GROUP (Consolidated Standards of Reporting Trials) (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, 134(8):663–694.
- [Anderson et al., 2016] Anderson, L., Oldridge, N., Thompson, D. R., Zwisler, A.-D., Rees, K., Martin, N., and Taylor, R. S. (2016). Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease Cochrane Systematic Review and Meta-Analysis. *Journal of the American College of Cardiology*, 67(1):1–12.
- [Barzi and Woodward, 2004] Barzi, F. and Woodward, M. (2004). Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology*, 160(1):34–45.
- [Bell et al., 2014] Bell, M. L., Fiero, M., Horton, N. J., and Hsu, C.-H. (2014). Handling missing data in RCTs; a review of the top medical journals. *BMC Medical Research Methodology*, 14:118.
- [Benzer, 2008] Benzer, W. (2008). Guidelines für die ambulante kardiologische Rehabilitation und Prävention in Österreich - Update 2008. Beschluss der Österreichischen Kardiologischen Gesellschaft vom Juni 2008. *Austrian Journal of Cardiology*, 15(9-10):298–309.
- [Fang et al., 2010] Fang, H., Johnson, C., Chevalier, N., Stopp, C., Wiebe, S., Wakschlag, L. S., and Espy, K. A. (2010). Using Propensity Score Modeling to Minimize the Influence of Confounding Risks Related to Prenatal Tobacco Exposure. *Nicotine & Tobacco Research*, 12(12):1211–1219.
- [Gripencrantz, 2014] Gripencrantz, S. (2014). *Evaluating the Use of Ridge Regression and Principal Components in Propensity Score Estimators under Multicollinearity*.
- [Gupta, 2011] Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, 2(3):109–112.
- [Hammill et al., 2010] Hammill, B. G., Curtis, L. H., Schulman, K. A., and Whellan, D. J. (2010). Relationship between cardiac rehabilitation and long-term risks of death and myocardial infarction among elderly Medicare beneficiaries. *Circulation*, 121(1):63–70.
- [Hanley and Foster, 2014] Hanley, J. A. and Foster, B. J. (2014). Avoiding blunders involving "immortal time". *International Journal of Epidemiology*, 43(3):949–961.
- [Harding, 2003] Harding, D. J. (2003). Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy. *American Journal of Sociology*, 109(3):676–719.
- [Harrell, 2001] Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer.
- [Hollis and Campbell, 1999] Hollis, S. and Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*, 319(7211):670–674.
- [Johnston and Guyatt, 2016] Johnston, B. C. and Guyatt, G. H. (2016). Best but oft-forgotten practices: Intention-to-treat, treatment adherence, and missing participant outcome data in the nutritional literature. *The American Journal of Clinical Nutrition*.
- [Krug, 2014] Krug, G. (2014). A simple algorithm to avoid misspecifying the propensity score equation. Evidence from a Monte Carlo simulation. *LASER discussion papers*, 83.
- [Lash and Cole, 2009] Lash, T. L. and Cole, S. R. (2009). Immortal Person-Time in Studies of Cancer Outcomes. *Journal of Clinical Oncology*, 27(23):e55–e56.
- [Lawler et al., 2011] Lawler, P. R., Fillion, K. B., and Eisenberg, M. J. (2011). Efficacy of exercise-based cardiac rehabilitation post-myocardial infarction: A systematic review and meta-analysis of randomized controlled trials. *American Heart Journal*, 162(4):571–584.e2.
- [Little and Kang, 2015] Little, R. and Kang, S. (2015). Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials. *Statistics in Medicine*, 34(16):2381–2390.
- [Members et al., 2012] Members, A. F., Perk, J., Backer, G. D., Gohlke, H., Graham, I., Reiner, e., Verschuren, M., Albus, C., Benlian, P., Boysen, G., Cifkova, R., Deaton, C., Ebrahim, S., Fisher, M., Germano, G., Hobbs, R., Hoes, A., Karadeniz, S., Mezzani, A., Prescott, E., Ryden, L., Scherer, M., SyvÄanne, M., Reimer, W. J. M. S. O., Vrints, C., Wood, D., Zamorano, J. L., Zannad, F., experts who contributed to parts of the Guidelines, O., Cooney, M. T., (cpg), E. C. f. P. G., Bax, J., Baumgartner, H., Ceconi, C., Dean, V., Deaton, C., Fagard, R., Funck-Brentano, C., Hasdai, D., Hoes, A., Kirchhof, P., Knuuti, J., Kolh, P., McDonagh, T., Moulin, C., Popescu, B. A., Reiner, e., Sechtem, U., Sirnes, P. A., Tendera, M.,

- Torbicki, A., Vahanian, A., Windecker, S., Reviewers, D., Funck-Brentano, C., Sirnes, P. A., Aboyans, V., Ezquerro, E. A., Baigent, C., Brotons, C., Burrell, G., Ceriello, A., Sutter, J. D., Deckers, J., Prato, S. D., Diener, H.-C., Fitzsimons, D., Fras, Z., Hambrecht, R., Jankowski, P., Keil, U., Kirby, M., Larsen, M. L., Mancia, G., Manolis, A. J., McMurray, J., Pajak, A., Parkhomenko, A., Rallidis, L., Rigo, F., Rocha, E., Ruilope, L. M., van der Velde, E., Vanuzzo, D., Viigimaa, M., Volpe, M., Wiklund, O., and Wolpert, C. (2012). European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). *European Heart Journal*, 33(13):1635–1701.
- [Piso et al., 2012]Piso, B., Tuechler, H., Kawalirek, S., Yazigi, M., and MÃ¼ller, R. (2012). Multizentrische, prospektive Kohortenstudie zur ambulanten kardiologischen Phase-III-Rehabilitation - Ein Studienprotokoll. *Journal of Cardiology*, 19(Pre Publishing Online).
- [Rosenbaum and Rubin, 1983]Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55. v-file.
- [Rosenbaum and Rubin, 1984]Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, 79(387):516–524.
- [Rubin, 2004]Rubin, D. B. (2004). On principles for modeling propensity scores in medical research. *Pharmacoepidemiology and Drug Safety*, 13(12):855–857.
- [Schulz et al., 2010]Schulz, K. F., Altman, D. G., Moher, D., and for the CONSORT Group (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, 340(mar23 1):c332–c332.
- [Suaya et al., 2009]Suaya, J. A., Stason, W. B., Ades, P. A., Normand, S.-L. T., and Shepard, D. S. (2009). Cardiac Rehabilitation and Survival in Older Coronary Patients. *Journal of the American College of Cardiology*, 54(1):25–33.
- [Weitzen et al., 2004]Weitzen, S., Lapane, K. L., Toledano, A. Y., Hume, A. L., and Mor, V. (2004). Principles for modeling propensity scores in medical research: A systematic literature review. *Pharmacoepidemiology and Drug Safety*, 13(12):841–853.
- [WMA, 2013]WMA (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20):2191.
- [Zhang, 2016]Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. *Annals of Translational Medicine*, 4(1).

Kommentierung des Endberichts
Ambulante kardiologische Rehabilitation
Teil VI: Multizentrische, prospektive
Kohortenstudie mit zwei Vergleichsgruppen
(mit/ohne Phase III)

http://eprints.hta.lbg.ac.at/1101/1/HTA-Projektbericht_Nr.90.pdf

Werner Pölz & Wolfgang Schimetta

1.) Hauptzielparameter-Auswertung:

Gegen die Umrechnung des ursprünglichen Kohortenunterschieds (Delta) von 0,29 Scorepunkten (kommt der als Studiengrundlage verwendeten Annahme von 0,32 Scorepunkten sehr nahe) in ein adjustiertes Delta mittels Propensity-Score-Stratifizierung (steht auch so im Prüfplan) ist nichts einzuwenden, allerdings ist die konkrete Umsetzung mit einem methodischen Fehler behaftet. Für die Propensity-Score-Variable „p0.A2.ErwerbStat.psHausfrau mann“ wird auf Seite 102 ein beta von 14,78 angegeben, was einer Odds Ratio von fast 3 Millionen entspricht. Das dazugehörige zweiseitige 95% Konfidenzintervall hat dann folglich Dimensionen, die i.d.R. von Null bis unendlich reichen. Dies ist ein bei logistischen Regressionen mit extremen Datenkonstellationen bekanntes Problem, das nicht nur sinnlose bzw. falsche Ergebnisse für die betreffende Variable induziert sondern auch die Ergebnisse der anderen Variablen maßgeblich verzerrt, also zu Pseudoergebnissen führt. Leider sind auch nicht von allen Variablen des Propensity Scores die Kohortenvergleiche verfügbar, diese wären in ihrer kompletten Zusammenstellung zu fordern (deskriptive und schließende Statistik), genauso wie eine Neuberechnung des Propensity Scores unter Ausschaltung des aufgetretenen Fehlers und mit Angabe der Odds Ratios incl. zweiseitigen 95% Konfidenzintervallen für alle einbezogenen Variablen. Somit kann der derzeit angeführte, aus der Adjustierung hervorgegangene Kohortenunterschied von 0,06 (mit dem zweiseitigen 95%igen Konfidenzintervall von -0,7 bis 0,6 – negatives Vorzeichen bedeutet Unterschied zugunsten der Interventionskohorte) nicht akzeptiert geschweige denn bewertet und diskutiert werden.

2.) Hauptzielparameter-Auswahl:

Die Generierung eines Summenscores für die Anzahl an Veränderungen zwischen suboptimal und optimal bei 6 Zielbereichen ist offenbar von den Kardiologen so akzeptiert worden,

obwohl die Trennschärfe nicht sehr groß ist: Es handelt sich um eine diskrete Variable mit 13 Zahlenwerten (von -6 bis +6), die meisten Patienten (>90%) haben aber nur einen Spielraum von 3 bis 4 Verbesserungen, wobei 3 der 6 Zielkategorien – „Blutdruck“, „Blutfette“ und „Blutzucker“ – medikamentös beeinflussbar bzw. optimierbar sind (da geht es zumeist eher um den Medikationsaufwand bis zur Zielerreichung als um die Zielerreichung per se) und auch die Zielkategorien „Rauchen“ und „Körpergewicht“ (mutmaßlich, ohne den Kardiologen vorgreifen zu wollen) nur indirekt mit einer ambulanten Rehabilitation zu tun haben, bleibt also nur „Leistungsfähigkeit“ als direkt beeinflussbare Zielkategorie. Dieses Manko betrifft vor allem Patienten aus der Interventionskohorte, die da auf Verbesserungen abzielt.

3.) Intent-To-Treat (ITT)-Analyse:

Durch die unerwartet hohe Anzahl an Drop Outs wurde post hoc beschlossen, keine echte ITT-Analyse (Primäranalyse) durchzuführen und sich nur auf Fälle zu beschränken, bei denen mindestens die Hälfte der Angaben zu den 12 Einzelzielen a priori vorhanden waren. Dies bedeutet, dass alle Patienten ohne Nachuntersuchung (T1) auch aus der ITT-Analyse herausfielen – eigentlich hätte ITT eine Fallzahl (n) von 157 bzw. 199 haben müssen, so wurde es ein n von 76 bzw. 81. Da die (meisten) Missing Values durch den Median der Realwerte der Kohorte ersetzt wurden, hätte ein komplettes bzw. korrektes ITT-n zwar vermutlich keinen größeren Kohortenunterschied erbracht, allerdings wahrscheinlich die Varianz massiv verringert.

4.) Per-Protocol (PP)-Analyse:

Es verwundert etwas, dass auf die übliche Durchführung einer begleitenden PP-Analyse komplett verzichtet wurde. Eine solche erscheint alleine schon wegen der offensichtlich extrem schwankenden Teilnahmefrequenz an der ambulanten Phase III Cardio-Reha empfehlenswert. So ist das Minimum an abgerechneten Reha-Trainingseinheiten 1 (!), und die Sozialversicherung bezahlt dann auch weniger, es ist also auch pekuniär nicht egal, ob Patienten kaum oder regelmäßig trainieren, was für die Berechtigung einer PP-Analyse spricht. Dazu müsste aber neben der Berücksichtigung der vorgegebenen Zeitfenster für die Aufnahme in die Studie und die Nachuntersuchung auch eine Mindestanzahl an absolvierten Trainings festgelegt werden (was offensichtlich a priori nicht erfolgt ist).

5.) Drop Outs:

Bei einer Studie mit dem offensichtlich hohen Aufwand und den hohen Kosten verwundert es etwas, dass es keine Möglichkeiten

gab, die Drop-Out-Rate von >55% (bezogen auf das final zugeordnete Kollektiv) auf erträgliche Größenordnungen zu senken (intensiverer Kontakt mit den Studienteilnehmern, Einbestellungen zur Nachuntersuchung mit Kostenübernahme und ggf. verbunden mit zulässigen Anreizen oder so...). So muss das Studienergebnis trotz der durchgeführten Prognosemodellierung, die eine Ergebnisrobustheit ergab, als potenziell verzerrt eingestuft werden (Attrition Bias). Das Nichterscheinen zur Nachuntersuchung kann nämlich auch andere Gründe haben als reinen Zufall (vgl. z.B. Verzerrungen bei Surveys durch Minder- oder Nichtteilnahme von Patienten in gewissen Outcome-Kategorien, die sich nicht durch eine Betrachtung der demografischen/anamnestischen Ausgangssituation zum Zeitpunkt der Fragebogenaussendung verlässlich vorhersagen lassen). Sollten etwa sowohl besonders gute als auch besonders schlechte Outcome-Kategorien von der vermehrten Nichtteilnahme betroffen gewesen sein, so würde die Vergleichbarkeit der Drop-Out-Anzahl in beiden Kohorten keine Garantie für eine kohortenunspezifische Verzerrung darstellen. Wenn es folglich zumindest teilweise einen systematischen Hintergrund (der zumindest 2 Varianten aufweist) haben kann, dass von mehr als der Hälfte der Studienteilnehmer die von Anfang an bekannten Follow-Up-Verpflichtungen negiert worden sind, dann ist ein deutliches Abweichen des Studienergebnisses von der Wahrheit keinesfalls auszuschließen und vielmehr konkret in Erwägung zu ziehen.

6.) Design:

Da der Hauptverband Österreichischer Sozialversicherungsträger als Auftraggeber fungierte und es offensichtlich dessen Ziel war, die Sinnhaftigkeit einer bisher finanzierten Therapie (ambulante Phase III Cardio-Reha) zu hinterfragen, ist zu diskutieren, ob bei dieser Konstellation nicht ein Nichtunterlegenheitsansatz (Nulltherapie nicht schlechter als ambulante Phase III Cardio-Reha) zu wählen gewesen wäre. Hätte der Hauptverband Österreichischer Sozialversicherungsträger die betroffenen Kardiologen (AGAKAR) aufgefordert, Daten zum Wirksamkeitsnachweis zu liefern, hätte die AGAKAR als Auftraggeber eine entsprechende Studie durchführen müssen, dann klar mit einem Überlegenheitsansatz (ambulante Phase III Cardio-Reha besser als Nulltherapie).

Ao.Univ.-Prof. Mag. Dr. Werner Pölz & Dr. Wolfgang Schimetta
Arbeitsgruppe zur Systemoptimierung klinischer Forschungsprojekte ASOKLIF
Abteilung für Angewandte Systemforschung und Statistik
Johannes Kepler Universität, Altenbergerstraße 69, A-4040 Linz
www.asoklif.at