

VOM BUNDESAMT FÜR GESUNDHEIT GESPONSERTE STUDIE

Längsschnittevaluation zur Genauigkeit und Vollständigkeit von Protokollen für klinische Studien – gibt es Belege für Verbesserungen?

Titel der Studie: Adherence to SPIrit REcommendations (ASPIRE)

AUTOREN: DMITRY GRYAZNOV, BENJAMIN KASENDA, ERIK VON ELM, MATTHIAS BRIEL

SCHLUSSBERICHT

ZUSAMMENFASSUNG

Background: Klar strukturierte und umfassende Studienprotokolle sind eine wesentliche Voraussetzung zur Gewährleistung der Sicherheit und des Wohlergehens von Versuchspersonen in klinischen Studien sowie für die Aussagekraft der Daten, eine erfolgreiche Durchführung der Studien und die Glaubwürdigkeit der Ergebnisse, namentlich im Fall von randomisierten klinischen Studien (randomized clinical trials, RCT). Unvollständige Protokolle gefährden alle Stufen des klinischen Forschungsprozesses und können schädliche Folgen für Patienten, Entscheidungsprozesse im Gesundheitswesen, die wissenschaftliche Gemeinschaft und die Gesellschaft als Ganzes haben. Finanzierungsstellen, Ehtikkommissionen, Regulierungsbehörden, medizinische Fachzeitschriften, Autoren systematischer Übersichtsarbeiten und anderen Gruppen verlassen sich auf Protokolle, wenn sie die Durchführung von Forschungsarbeiten und Berichte darüber bewerten. Evidenzbasierte Daten zu Kohorten von RCT-Protokollen, die in den 1990er-Jahren von Ethikkommissionen genehmigt worden waren, zeigen sehr grosse Unterschiede hinsichtlich Inhalt und Qualität dieser Protokolle. Als Reaktion auf diese Problematik wurden 2013 Richtlinien veröffentlicht: die «Standard Protocol Items: Recommendations for Interventional Trials» (SPIRIT). Diese sollen zur Verbesserung der Genauigkeit und Vollständigkeit von klinischen Versuchsprotokollen beitragen. Sie bestehen aus evidenzbasierten Empfehlungen mit einer Checkliste von 33 wichtigen Punkten, die in Protokollen im Sinne eines Mindeststandards zu beachten sind. Im Januar 2014 traten das Humanforschungsgesetz und die dazugehörigen Verordnungen in Kraft, die unter anderem eine Neuverteilung der Rollen und Abläufe für die Schweizer Ethikkommissionen und Swissmedic vorsehen. In diesem Kontext wurden neue Richtlinien für die Ethikkommissionen und die Prüfer entwickelt, die sich auf den SPIRIT-Rahmen stützen. Bei der Einführung der neuen Schweizer Gesetzgebung wurde eine Evaluation vorgesehen. Da eine Optimierung der Forschungsqualität zu den Zielen des Humanforschungsgesetzes gehört, sponserte das Bundesamt für Gesundheit dieses Evaluationsprojekt, um in Erfahrung zu bringen, wie genau und vollständig die von Schweizer Ethikkommissionen genehmigten RCT-Protokolle vor bzw. nach der Einführung des Humanforschungsgesetz (HFG) sind.

Ziele: Als Messgrösse für die Qualität der klinischen Forschung im Allgemeinen wählten wir die Genauigkeit/Vollständigkeit der Protokolle randomisierter klinischer Studien (RCT-Protokolle). Wir konzentrierten uns auf RCT, weil deren Ergebnisse in der Regel eher in die klinische Praxis und in Richtlinien einfließen als Beobachtungsstudien. Angesichts der Risiken und des Aufwands der Patienten bei RCT-Studien erfordert eine umsichtige und professionelle Durchführung solcher Studien zudem von den Prüfern ein besonders hohes Mass an Verantwortung. Wir definierten für dieses Projekt folgende Ziele:

1. Analyse der Genauigkeit und Vollständigkeit von RCT-Protokollen, die von Schweizer Ethikkommissionen vor bzw. nach der Einführung des HFG in der Schweiz im Januar 2014 genehmigt wurden (2012 bzw. 2016), anhand der SPIRIT-Checkliste.
2. Evaluation des Anteils von registrierten und insbesondere prospektiv registrierten Protokollen (Messung der Transparenz) vor der Einführung des HFG im Januar 2014 und danach (in nationalen oder internationalen Registern).
3. Bestimmung von Merkmalen einer Studie, die mit der Nichtbeachtung der Punkte der SPIRIT-Checkliste verbunden sind, unter anderem ein möglicher Zusammenhang zwischen dem Genehmigungsjahr (2012 bzw. 2016) und der Art des Sponsors (Industrie bzw. andere) sowie dem Genehmigungsjahr (2012 bzw. 2016) und der deklarierten methodologischen Unterstützung durch Clinical Trial Units (CTU) oder Clinical Research Organisations (CRO) (ja/nein).
4. Analyse der Genauigkeit und Vollständigkeit von Schweizer RCT-Protokollen im Vergleich zu nicht-schweizerischen RCT-Protokollen (die in Deutschland oder Kanada 2012 oder 2016 genehmigt wurden) anhand der SPIRIT-Checkliste. Ausserdem wurde der Anteil von registrierten schweizerischen bzw. nicht-schweizerischen Protokollen analysiert.

Methoden: Für diese Vorher-Nachher-Studie konnten wir die Unterstützung und die bewährte Zusammenarbeit mit swissethics, den sieben Schweizer Ethikkommissionen und zwei ausländischen Ethikkommissionen (Universitätsklinikum Freiburg, Deutschland und Hamilton Health Sciences, Kanada) in Anspruch nehmen. Berücksichtigt wurden alle 2012 und 2016 in der Schweiz genehmigten RCT-Protokolle, bei denen gesundheitliche Outcomes von Behandlungen untersucht wurden. Nicht berücksichtigt wurden RCT über verschiedene Verabreichungsdosen oder -wege desselben

Arzneimittels (Dosisfindungsstudien) sowie als Pilot- oder Machbarkeitsstudien bezeichnete Versuche. Ebenfalls ausgeschlossen wurden Studien, bei denen es hauptsächlich um Evaluationen der Pharmakokinetik oder der Physiologie (Studien mit gesunden Freiwilligen) oder um gesundheitsökonomische Aspekte ging, sowie Studien mit Tierversuchen oder Gewebebanken und quasirandomisierte Studien. Für jedes Protokoll wurden zuerst Daten über die allgemeinen Merkmale der Studie erhoben (z.B. geplante Stichprobengrösse, einzel- oder multizentrisch). In einem zweiten Teil analysierten wir, ob mit dem RCT-Protokoll Informationen zu den einzelnen Punkten der SPIRIT-Checkliste geliefert wurden (ja / nein). Bei den 2012 genehmigten Protokollen erfolgte die Datenextraktion in Teams von zwei Reviewern mit methodologischer Schulung, die unabhängig voneinander arbeiteten und danach ihre Ergebnisse und allfällige Abweichungen diskutierten. Für die 2016 genehmigten Protokolle wurden aus Machbarkeitsgründen lediglich 30% der eingeschlossenen RCT-Protokolle unabhängig und doppelt analysiert. Unser wichtigstes Outcome war die Einhaltung der einzelnen Punkte der SPIRIT-Checkliste. Wir berechneten diese Einhaltung als Anteil der in den Versuchsprotokollen beachteten Punkte der SPIRIT-Checkliste, ausgedrückt als Durchschnitts- und Medianwert der vom Protokoll beachteten Punkte. Wir verwendeten multivariate Regressionsanalysen, um zu bestimmen, ob ein Zusammenhang bestand zwischen der Nichteinhaltung der SPIRIT-Checkliste und (i) dem Jahr der Genehmigung (2012/2016), (ii) der Art des Sponsorings (Industrie/Andere), (iii) der geplanten Stichprobengrösse (kontinuierliche Variable), (iv) der Zentrumsart (einzel-/multizentrisch), (v) der angegebenen methodologischen Unterstützung durch CTUs oder CROs (ja/nein) und der Behandlungsart (mit/ohne Arzneimittel). Zur Prüfung unserer Hypothese einer deutlicheren Verbesserung bei Protokollen von Studien, die nicht von der Industrie gesponsert waren, bezogen wir die entsprechende Interaktion (Jahr der Genehmigung * Art des Sponsors) in unser Regressionsmodell mit ein. Denselben Ansatz verwendeten wir, um unsere Hypothese zu testen, dass der Anteil der eingehaltenen SPIRIT-Punkte in Protokollen mit methodologischer Unterstützung (durch CTU oder CRO) weniger deutlich verbessert wurde als in solchen ohne diese Unterstützung. Für den Vergleich mit nichtschweizerischen RCT-Protokollen schlossen wir in unser Regressionsmodell eine zusätzliche unabhängige Variable ein (schweizerisches/nichtschweizerisches Protokoll). Wir stützten uns dabei auf alle verfügbaren Daten aus Protokollen, die 2012 oder 2016 in der Schweiz, in Freiburg oder in Hamilton genehmigt wurden. Zusätzlich nahmen wir eine Analyse der Interaktion wie oben beschrieben vor, wobei wir ausschliesslich nichtschweizerische RCT-Protokolle berücksichtigten.

Zur Schätzung des Anteils von prospektiv registrierten RCT-Protokollen in den Jahren 2012 und 2016 als Messgrösse für Transparenz durchsuchten wir alle RCT-Protokolle und andere an die Ethikkommissionen eingereichten Unterlagen auf eine Registrierungsnummer sowie die WHO International Clinical Trials Registry Platform, clinicaltrials.gov und Google Scholar nach entsprechenden Registrierungsdaten. Dann verglichen wir die entsprechenden Registrierungsdaten mit dem Aufnahmedatum des ersten Teilnehmenden. Falls das Registrierungsdatum vor oder innerhalb eines Monats nach dem Aufnahmedatum des ersten Teilnehmenden lag, erachteten wir die RCT als prospektiv registriert.

Ergebnisse: Wir analysierten 400 von Schweizer Ethikkommissionen bewilligte RCT-Protokolle (von denen 183 im Jahr 2012 und 217 im Jahr 2016 genehmigt wurden). Insgesamt fanden wir keine Unterschiede beim Anteil der im Protokoll eingehaltenen SPIRIT-Punkte zwischen RCT-Protokollen im Jahr 2012 (Median 74%, Interquartilabstand [IQR] 64%-80%) und 2016 (Median 76%, IQR 69%-82%). Eine deutliche Verbesserung war hingegen in der Teilgruppe der nicht von der Industrie gesponserten Protokolle festzustellen (d.h. von Prüfern initiierte RCT): Der Median stieg von 65% (IQR 56%-74%) im Jahr 2012 auf 76% (IQR 66%-83%) im Jahr 2016. Konkret wurde die Einhaltung einer breiten Auswahl von SPIRIT-Punkten und -Teilpunkten verbessert: Bei 23 verschiedenen Punkten war eine Verbesserung um mindestens 10% beim Anteil der Protokolle festzustellen, die Angaben zum entsprechenden Punkt machten. Bei den von der Industrie gesponserten Protokollen ist der Anteil der angegebenen SPIRIT-Punkte über den betrachteten Zeitraum stabil geblieben (Median 79%, IQR 75%-82% im Jahr 2012 gegenüber Median 77%, IQR 72%-82% im Jahr 2016). Dieser Teilgruppen-Effekt war unabhängig von der geplanten Grösse der RCT, von der Unterstützung durch eine CRU oder CRO, von der Behandlung (mit/ohne Arzneimittel) und von der Zentrumsart (einzel-/multizentrisch). Wir

stellten fest, dass die nachfolgenden RCT-Merkmale signifikant und unabhängig voneinander mit einer niedrigeren SPIRIT-Einhaltung korrelierten: Einzelzentrum, keine angegebene Unterstützung durch CTU oder CRO, nicht von der Industrie gesponsert (d.h. durch Prüfer initiiert) und 2012 genehmigt. Wir beobachteten keinen Teilgruppen-Effekt für Protokolle mit bzw. ohne methodologische Unterstützung durch CTU oder CRO. Die Analyse von RCT-Protokollen, die ausserhalb der Schweiz genehmigt wurden (Freiburg in Deutschland und Hamilton in Kanada: 79 Protokolle 2012 und 76 Protokolle 2016) ergab einen ähnlichen Teilgruppen-Effekt mit einer leichten Verbesserung bei den nicht von der Industrie gesponserten Protokollen im Jahr 2016 gegenüber 2012 (Median 59%, IQR 53%-69% im Jahr 2012 gegenüber Median 62%, IQR 54%-70% im Jahr 2016), während der Anteil der angegebenen SPIRIT-Punkte für die von der Industrie gesponserten Protokolle ungefähr gleich blieb. Zudem war tendenziell eine deutlichere Verbesserung bei den in der Schweiz genehmigten Protokollen gegenüber nichtschweizerischen RCT festzustellen, dieser Unterschied war aber nicht statistisch signifikant.

In beiden Jahren wurden industriegesponserte Protokolle häufiger registriert und häufiger prospektiv registriert als nicht von der Industrie gesponserte Protokolle, sowohl in der Schweiz als auch im Ausland (2012: Anteil prospektiv registrierter Protokolle 89%, 95%-CI, 82%-94% für industriegesponserte Schweizer Protokolle bzw. 69%, 95%-CI, 58%-77% für nicht von der Industrie gesponserte Schweizer Protokolle). Es gab jedoch keine Hinweise darauf, dass der Anteil bei in der Schweiz oder im Ausland registrierten bzw. prospektiv registrierten Protokolle stieg, abgesehen von einer Ausnahme: Der Anteil bei prospektiv registrierten Protokollen unter den nicht von der Industrie gesponserten Protokollen erhöhte sich ausserhalb der Schweiz von 73% (95%-CI, 56%-85%) im Jahr 2012 auf 83% (95%-CI, 67%-92%) im Jahr 2016.

Schlussfolgerungen: Diese Vorher-Nachher-Studie weist darauf hin, dass die Vollständigkeit der Protokolle von in der Schweiz genehmigten, nicht von der Industrie gesponserten RCT zwischen 2012 und 2016 leicht besser wurde (2012: 65% der SPIRIT-Punkte erfüllt; 2016: 76% der SPIRIT-Punkte erfüllt; Medianwerte). Verglichen mit Protokollen, die ausserhalb der Schweiz genehmigt wurden, scheint die Verbesserung bei den nicht von der Industrie gesponserten Schweizer Protokollen ausgeprägter, der Unterschied ist aber nicht statistisch signifikant. Es gab dagegen keine Anhaltspunkte für eine Verbesserung des Anteils bei in der Schweiz registrierten oder prospektiv registrierten Protokollen.

Dies weist auf einen internationalen Trend hin zu einer leichten Verbesserung bei der Vollständigkeit von nicht von der Industrie gesponserten (von Prüfern initiierten) RCT-Protokollen hin. Verantwortlich dafür sind wahrscheinlich mehrere Gründe, unter anderem die Veröffentlichung der SPIRIT-Richtlinien 2013 und ihre Förderung durch Forschungseinrichtungen, Finanzierungsstellen und medizinische Fachzeitschriften; die laufende Diskussion über die Bedeutung der Publikation von Protokollen, eine durchdachte Planung von RCT und die Vermeidung von Publikationsverzerrungen (Reporting Bias) in der wissenschaftlichen Gemeinde; und schliesslich eine bessere Ausbildung von klinisch tätigen Wissenschaftlern im Bereich der RCT-Methodologie. Gewisse Ergebnisse weisen jedoch darauf hin, dass die Verbesserung bei den nicht von der Industrie gesponserten RCT-Protokollen, die in der Schweiz bewilligt wurden, deutlicher ausfiel als bei den ausserhalb der Schweiz genehmigten Protokollen, was darauf schliessen lässt, dass das HFG möglicherweise einen zusätzliche Wirkung hatte, vermutlich durch Anleitungen und Formulare für Studienprotokolle von swissethics, die von den Forschenden ausdrücklich begrüsst wurden. Bezüglich Transparenz braucht es weitere Anstrengungen zur Umsetzung der prospektiven Registrierung von RCT-Protokollen in der Schweiz, insbesondere bei Protokollen, die nicht von der Industrie gesponsert sind. Am besten lässt sich dieses Ziel vermutlich erreichen durch gemeinsame, abgestimmte Massnahmen verschiedener Akteure der klinischen Forschung in der Schweiz.