

Checkliste für Qualitätssicherung bei der Planung, Durchführung und Analyse von Diagnosestudien in der klinischen Chemie

Keller T

ACOMED statistik, Leipzig, Deutschland
info@acomed-statistik.de

Einleitung und Fragestellung Die Evaluierung der diagnostischen Güte von Diagnoseverfahren ist in verschiedenen Richtlinien geregelt [1, 2]. In der Realität sind jedoch vielfältige Verletzungen der Standards anzutreffen [3-5]. Die unzureichende Planung und Durchführung von Diagnosestudien führen häufig zu fehlerhaften Studienergebnissen, wobei sowohl Überschätzung als auch Unterschätzung der diagnostischen Güte die Folge sein können. Die mangelhafte Charakterisierung der Laboreigenschaften des zu untersuchenden Tests im Vorfeld der Diagnosestudie kann weiterhin zu nicht verwertbaren Ergebnissen der Studie führen. Schließlich sind die statistischen Analysen so zu planen und durchzuführen, dass die Ergebnisse der Studie von klinischer Relevanz sind. Basierend auf mehrjährigen Erfahrungen in der Planung, Durchführung und Analyse von Diagnosestudien wurde eine Checkliste entwickelt, die sich – anders als viele Richtlinien - an praktischen Fragestellungen orientiert, und insbesondere zusätzlich die Messqualität einbezieht. Sie beschreibt für Diagnosestudien spezifische, gegenüber üblichen klinischen Studien z.T. hinausgehende Anforderungen. Hierbei wurden Empfehlungen anderer Richtlinien berücksichtigt, z.B. jene für die Laboranalytik bei Bioäquivalenzuntersuchungen [6]. Mit Hilfe der Checkliste sollen die Auswertbarkeit der Studie sowie Richtigkeit, Wiederholbarkeit und Nachvollziehbarkeit des Studienergebnisses sichergestellt werden.

Ergebnis: Checkliste Studienplanung

Kriterium	Kommentar, Fehler, Konsequenz bei Nichterfüllung
Nennung der beabsichtigten Gebrauchs (Intended use) des Diagnostikums, Rolle des Tests in Entscheidungsfindung, Studienpopulation	Die Kenntnis der beabsichtigten Gebrauchs sowie des diagnostischen Pfades sind ausschlaggebend für das Studienziel sowie die Ein- und Ausschlusskriterien. Konsequenzen bei Nichtberücksichtigung können Festlegung eines nichtadäquaten Studienziels bzw. ein Studienergebnis ohne klinische Relevanz sein. Mitunter wird auch die Studienpopulation durch Gesunde (z.B. Laborpersonal), oder Patienten mit unerheblichen Erkrankungen etc. aufgefüllt.
Festlegung der zu erreichenden diagnostischen Güte	Die zu erreichende diagnostische Güte ergibt sich z.B. aus Erkrankung, therapeutischer Konsequenz, konkurrierenden Tests. Die Aufdeckung eines Gruppenunterschiedes kann kein Ziel einer Diagnosestudie sein. Die Fallzahlberechnung auf der Basis des Nachweises eines Gruppenunterschiedes führt zu zu geringen Fallzahlen.
Angabe eines klinisch sinnvollen Sensitivitäts- bzw. Spezifitätsbereiches	AUC der ROC-Analysen beurteilen den Bereich aller Falsch-Positiv-Raten. Klinisch interessant sind aber nur ausgewählte Bereiche, z.B. hoher Spezifität. Bei Nichtberücksichtigung können klinisch nicht relevante Ergebnisse erzielt werden.
Verwendung anerkannter Größen zur Beschreibung der diagnostischen Güte	Dies sind Sensitivität, Spezifität, Vorhersagewerte, Ergebnisse ROC Analyse. Methode der Wahl wären partielle ROC-Kurven, die sich aber (noch) nicht durchgesetzt haben. Verwendung von Sensitivität und Spezifität bei minimaler Falsch-Negativrate, Youden-Index und Effizienz sind meist klinisch nicht relevant und werden nicht empfohlen. Verwendung anderer Größen lässt Schlussfolgerung zu, dass bzgl. o.g. Parameter das Studienergebnis nicht der Hypothese entspricht.
Festlegung des Studientypes (meist Phase II oder III) nach GMDS-Memorandum oder entsprechend	Nach [7] überschätzt Studientyp Phase II (Einschlusskriterium ist bekannte Erkrankung) das Odds-Ratio ggü. Studientyp Phase III (Einschlusskriterium ist Erkrankungsverdacht) um Faktor 3. Ein Studienergebnis bei Typ II muss entsprechend bewertet werden (ggf. Überschätzung der diagnostischen Güte ggü. klinischer Anwendungssituation).
Stellungnahme zum Goldstandard	Unzureichende Goldstandards sind eine der großen Schwierigkeiten bei Diagnosestudien. Studienergebnis verzerrt.
Stellungnahme zu wesentlichen Verzerrungen	Verifikations-Verzerrung (ungleiche Anwendung Goldstandard für Gruppen), Einschluss des Testresultats (bzw. eines ähnlichen Testverfahrens) in Goldstandard, Selektions- und Spektrums-Verzerrungen (Patientenpopulation nicht repräsentativ), ggf. auch Review-Verzerrung (Kenntnis des Goldstandards bei Test-Durchführung), Verzerrung infolge Messreihenfolge (bei Vergleich von Tests: Kenntnis des Resultat des anderen Tests).
Festlegung des Probenahmezeitpunktes	Zeitpunkt Probenahme sollte klinischem Einsatz des Diagnostikums entsprechen. Auf Einheitlichkeit achten.
Verwendung alter Proben	Bei Verwendung von Proben aus Probenbank: Konsequenzen bzgl. Repräsentativität, Präanalytik (s.u.) darstellen.

Ergebnis: Checkliste Präanalytik/ Qualitätssicherung Laboranalytik [5]

Beschreibung Probenahme, Lagerung	Wichtig bzgl. Blutentnahmesysteme und Abstrichen, ggf. Identifizierung eines Einflussfaktors, Einfluss Lagerdauer
Abschätzung Präzision (intraserieller, interserieller VK) vor Diagnostikstudie	Studientyp I nach GMDS-Memorandum sollte vor Diagnostikstudie vorliegen Diagnosestudie ohne greifbare Resultate, wenn $VK > \text{Effekt}$
Feststellung folgender weiterer Laboranalytischer Kenngrößen der Methode (z.B. aus Literatur, vom Hersteller, Begleitzettel IVD-Produkt). Konzentrationsbereich und Linearität. Detektionsgrenze (LOD), Bestimmungsgrenze (LOQ), Spezifität ggü. Matrixeffekten, Richtigkeit (Recovery), Analytenstabilität, Stabilität des Messgerätes bzw. des Assays. Zu jedem dieser Punkte sollte ein Kommentar erstellt werden, und Auswirkungen bzgl. des Ergebnisses diskutiert werden.	
Messtägliche bestimmte Kontrollproben	Hinweis im Report auf Ergebnisse Kontrollproben. Ggf. Auswirkungen auf Studienergebnis feststellen und diskutieren.
Poolproben	Im Report sind die Ergebnisse der in den Messserien mitlaufenden Poolproben aufzuführen (Schätzer für Richtigkeit, Präzision, Analyten- und Messgeräte-stabilität)

Ergebnis: Checkliste Monitoring

Durchführung und Inhalte des Monitoring planen und ankündigen	Monitoring in den Klinischen Labors mit Sichtung und Hinterfragung der Qualitätsmaßnahmen ist zu empfehlen. Monitoring in Kliniken : Schwerpunkt Referenzstandard, Zeitpunkt Probenahme, Erfüllung Einschlusskriterien bei Vergleichsgruppe
---	---

Ergebnis: Checkliste Datenmanagement Labordaten, Datenvalidierung

Planung der Mess-Serien: Serien vs. Messung bei Probeneingang	Da intraserieller $VK < \text{interserieller VK}$, unterschiedliche Studienergebnisse. Gemeinsame bzw. zufällig verteilte Messungen an Proben von Kranken und Nichterkrankten ist obligat. Konsequenz: ein zu optimistisches oder verzerrtes Studienergebnis.
Festlegung der Einheiten	wichtig bei multizentrischen Studien (z.B. Massekonzentrationen vs. Stoffmengenkonzentrationen)
Kriterien für Messwiederholungen festlegen	Bei z.B. Doppelmessungen sollte prospektiv festgelegt werden, welche Abweichungen zulässig sind, was bei Abweichungen passiert (z.B. Messwiederholung) und wie die Messergebnisse zusätzlicher

	Wiederholungsmessungen verwendet werden.
Planung des Umgangs mit Ausreißern, fehlenden Werten, Messwerten außerhalb des Messbereiches, unerwarteten Werten, Festlegung des Datenflusses. Identifikation von Punkten mit Fehlerpotential. Datenvalidierung	Entfernen von Ausreißern ist laut Richtlinien erlaubt, sofern Begründung vorliegt. Zu empfehlen ist eine prospektive Festlegung der Kriterien. – Fehlende Werte können bestimmte Subpopulationen gehäuft betreffen (Grauzone, frühe Stadien). - Typisch sind folgende Fehler bzw. Schwierigkeiten: Angabe als String (z.B. „<1“; „>1000“), fehlendes bzw. falsches Dezimaltrennzeichen, - Die Verwendung von Tabellenbearbeitungsprogrammen zzgl. Datentransfer mittels Kopieren und Einfügen sind üblich und fehleranfällig.
Ergebnis: Checkliste Auswertung	
Planung zur Transformation der Daten	Wegen häufig notwendiger Logtransformation: Umgang mit Messwerten = 0 festlegen (z.B. = untere Nachweisgrenze).
Analyse der Einflussfaktoren (zusätzlich zu klinischen Merkmalen): Zentrum, Messtag, Messreihenfolge, Lagerdauer, Messserien, ggf. Chargen, Laborant	
Prospektive Planung von Subgruppenanalysen, Bei Vergleichen: Subgruppenauswahl so, dass fair für beide Tests	
Weitere Aspekte: vgl. Planung	

Tab. 1: Liste von Maßnahmen, die in bei Planung, Durchführung und Analyse einer Diagnosestudie einbezogen werden sollten

Diskussion

Die Checkliste erfüllt nicht die Maßstäbe einer Guideline [8], kann aber hilfreich in der Praxis sein, denn die in Tab. 1 angegebenen Maßnahmen sind zum großen Teil einfach umsetzbar. Häufig bedarf es nur einer Kommentierung, dass der Sachverhalt bei Studienplanung/-durchführung berücksichtigt wurde. Bereits der Einsatz der Liste im Sinne einer Checkliste hilft, Diagnosestudien so zu gestalten, dass sie richtige und verwertbare Ergebnisse erbringen. Der Einsatz der Checkliste führt aber auch dazu, dass Studien, die mit hoher Wahrscheinlichkeit zu keinem oder einem unrichtigen Ergebnis führen können, gar nicht erst durchgeführt werden. Die Tabelle ist ein Auszug der entsprechenden internen SOP im Unternehmen des Autors. - Die Liste erhebt keinen Anspruch auf Vollständigkeit, sondern ist als Diskussionsgrundlage gedacht. Wichtig ist, bei der Festlegung der Maßnahmen darauf zu achten, dass sie nicht zu allgemeiner Natur sind, dass sie umsetzbar sind, dass sie nutzerfreundlich angelegt sind, und dass sich zusätzlich entstehende Kosten und Aufwand in engen Grenzen halten.

Literatur

- [1] GMDs: Köbberling J, Trampisch H-J, Windeler J: Memorandum zur Evaluierung diagnostischer Maßnahmen Schattauer 1989, Stuttgart, New York
- [2] NCCLS. Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots; Approved Guideline. NCCLS Document GP10-A (ISBN 1-56238-285-3). NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087, 1995.
- [3] Reid MC, Lachs MS, Feinstein AR: Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA. 1995;274:645-651
- [4] Rutjes AWS, Reitsma RB, Di Nisio M, Smidt N, van Rijn JC and Bossuyt PMM: Evidence of bias and variation in diagnostic accuracy studies. Can Med Assoc J 2006;174:469-476
- [5] Obuchowski NA, Lieber ML, Wians Wians FH. ROC Curves in Clinical Chemistry: Uses, Misuses, and Possible Solutions. Clin Chem 2004;50:1118-1125
- [6] FDA Guidance for Industry - Bioequivalence guidance (http://www.fda.gov/cvm/Guidance/bioequivalence_Oct02.pdf) 2002.
- [7] Lijmer JG, Mol BW, Heisterkamp S; Bonsel GJ, Prins MH, Meulen JHP, Bossuyt PMM. Empirical evidence of design related bias in studies of diagnostic tests, JAMA 1999, 282: 1061-6
- [8] Oosterhuis WP, Bruns DE, Watine J, Sandberg S, Horvath AR. Evidence-Based Guidelines in Laboratory Medicine: Principles and Methods. Clin Chem 2004;50:806-818