

Aus dem Bereich Innere Medizin

Klinik für Hämatologie, Internistische Onkologie und Diabetologie

Westfalz-Klinikum Kaiserslautern, Akademisches Lehrkrankenhaus der Johannes-Gutenberg-Universität Mainz

Direktor: Prof. Dr. med. Gerhard Held

## **Übereinstimmung zwischen Beurteilern bei der Bewertung der Kausalität und Erwartbarkeit in klinischen Studien**

Dissertation zur Erlangung des Grades eines

Doktors der Medizin

der Medizinischen Fakultät der UNIVERSITÄT DES SAARLANDES

2022

vorgelegt von Giorgi Natchkebia,  
geboren am 09.10.1993 in Tbilissi

# Inhaltsverzeichnis

Inhaltsverzeichnis .....	II
Abkürzungsverzeichnis .....	IV
Abbildungs- und Tabellenverzeichnis.....	VI
1. Zusammenfassung und Abstract.....	1
1.1 Zusammenfassung .....	1
1.2 Abstract .....	3
2. Einleitung.....	5
2.1 Arzneimittelsicherheit in Klinischen Studien .....	5
2.1.1 Unerwünschte Arzneimittelwirkung .....	7
2.1.2 Unerwünschtes Ereignis .....	8
2.1.3 Schwerwiegendes unerwünschtes Ereignis .....	9
2.1.4 Bestimmung der Erwartbarkeit.....	10
2.1.5 Dokumentation und Meldung .....	13
2.2 Methoden der Kausalitätsbewertung.....	14
2.2.1 Experteneinschätzung .....	15
2.2.2 WHO-UMC Skala.....	18
2.2.3 Algorithmische Methoden.....	21
2.2.4 Probabilistische Methoden.....	24
2.3 Kausalitätsbewertung in klinischen Studien .....	24
2.3.1 Kausalitätszuordnung durch Lokale Studienärzte .....	26
2.3.2 Zweitbewertung der schwerwiegenden unerwünschten Ereignisse durch den Sponsor. ....	28
2.4 Rezidivtherapie des aggressiven B- und T-Zell-Lymphoms.....	29
2.4.1 Patienten geeignet für eine Hochdosistherapie mit autologer Blutstammzelltransplantation .....	29
2.4.2 Patienten nicht geeignet für eine Hochdosistherapie mit autologer Blutstammzelltransplantation .....	30
2.4.3 R-GemOx .....	30
3 Fragestellung und Zielsetzung .....	31
4 Material und Methoden .....	33
4.1 Synopse der NIVEAU Studie .....	33
4.2 Flussdiagramm der NIVEAU Studie .....	37
4.3 Ein- und Ausschlusskriterien der NIVEAU Studie .....	38
4.4 Dokumentation und Meldung .....	39
4.5 Evaluation der Kausalität und Erwartbarkeit durch Rater .....	42

4.6	Statistische Methoden.....	43
5	Ergebnisse.....	44
5.1	Beteiligte Zentren.....	44
5.2	Beschreibung des Patientenkollektivs.....	45
5.3	Charakterisierung der Schwerwiegenden unerwünschten Ereignisse .....	46
5.4	Qualität der Meldungen.....	50
5.5	Kausalitätsprüfung der schwerwiegenden unerwünschten Ereignisse .....	50
5.6	Die Evaluation der Erwartbarkeit der schwerwiegenden unerwünschten Ereignisse.....	54
6	Diskussion .....	55
6.1	Kausalitätsprüfung der schwerwiegenden unerwünschten Ereignisse .....	55
6.2	Möglichkeiten zur Besserung der Übereinstimmung zwischen lokalen Studienärzten und Sponsor.....	61
6.3	Die Evaluation der Erwartbarkeit der schwerwiegenden unerwünschten Ereignisse.....	63
7	Literaturverzeichnis.....	67
8	Publikationen .....	74
9	Anhang .....	75
9.1	Beurteilungsbogen aus der NIVEAU Studie .....	75
9.2	Fallberichtsbogen aus der NIVEAU Studie.....	76
10	Dankausgang .....	79
11	Lebenslauf.....	80

## Abkürzungsverzeichnis

Abkürzung	Erläuterung
AESI	Adverse Events of special interest
ASZT	Allogene Stammzelltransplantation
BfarM	Das Bundesinstitut für Arzneimittel und Medizinprodukte
CCS	Canadian Cardiovascular Society
CRF	Case Report Form
CTC-AE	Common Terminology Criteria for Adverse Events
eCRF	Elektronisches Case Report Form
ECOG	Eastern Cooperative Oncology Group
EU	Europäische Union
FDA	Lebensmittelüberwachungs- und Arzneimittelbehörde der Vereinigten Staaten
GCP	Good Clinical Practice
GCP-V	GCP-Verordnung
HCT-CI	Hematopoietic Cell Transplantation-specific Comorbidity Index
HIV	Human Immunodeficiency Virus
IB	Investigator's Brochure
ICH	International Council for Harmonisation
MTD	Maximal tolerierbare Dosis
mg	Milligramm
n.n.b.	Nicht näher bezeichnet

NYHA	New York Heart Association
PEI	Das Paul-Ehrlich-Institut
RSI	Referenzsicherheitsinformation
(R)-DHAP	Rituximab, Dexamethason, hochdosiertes Ara-C und Cisplatin
R-GemOx	Rituximab, Gemcitabin, Oxaliplatin
R-ICE	Rituximab, Ifosfamid, Carboplatin, Etoposid
SUAW	Schwerwiegende unerwünschte Arzneimittelwirkung
SUE	Schwerwiegendes unerwünschtes Ereignis
SUSAR	Verdachtsfall einer unerwarteten, schwerwiegenden Nebenwirkung
UAW	Unerwünschte Arzneimittelwirkung
UE	Unerwünschtes Ereignis
ULN	Upper Limit Of Normal
UMC	Uppsala Monitoring Center
ZKS	Zentrum für Klinische Studien
µl	Mikroliter

## Abbildungs- und Tabellenverzeichnis

Abbildung/Tabelle	Bezeichnung	Seite
Abbildung 1	Fortlaufende und systematische Überwachung während des gesamten Lebenszyklus des Medikaments.	5
Abbildung 2	Zusammenhänge zwischen verschiedenen Ereignissen in der Pharmakovigilanz	13
Abbildung 3	Das Flussdiagramm zur Kausalitätsbewertung in klinischen Studien, präsentiert von Ping-Hwa-Su und Ralph W. Stoll.	23
Abbildung 4	Der übliche Vorgang der Kausalitätsbewertung durch Studienärzte nach Mukherjee et al.	28
Abbildung 5	Therapie der Patienten eingeschlossen in den experimentellen und Standardarm der NIVEAU-Studie.	37
Abbildung 6	Die Abbildung zeigt die SUEs, die die Beurteiler A und B hinsichtlich der Erwartbarkeit bewerteten. Zur Bestimmung der Intrarater-Reliabilität bewertete der Beurteiler B zusätzlich 117 SUEs zum zweiten Zeitpunkt im Abstand von mindestens 3 Monaten.	42
Abbildung 7	Das Balkendiagramm zeigt, in welcher Phase bei den Patienten aus dem experimentellen Arm der Studie die SUEs auftraten.	47
Abbildung 8	Gründe für die Einstufung der SUEs aus den beiden Studienarmen als schwerwiegend durch lokale Studienärzte.	49
Tabelle 1	Abstufung der Kausalitätskategorien durch PEI, BfarM und WHO.	15
Tabelle 2	Die Kausalitätsbewertungsskala von WHO-UMC.	19
Tabelle 3	Der Naranjo-Fragebogen und dessen Punkteverteilung.	22
Tabelle 4	Synopse der NIVEAU-Studie.	36
Tabelle 5	Zulassungskriterien zu der NIVEAU-Studie.	38
Tabelle 6	Ausschlusskriterien der NIVEAU-Studie.	39
Tabelle 7	Qualitative Bewertungen der Kappa-Werte sowie der Korrelationskoeffizienten nach Spearman im medizinischen Bereich.	44

Tabelle 8	Herkunftsländer der unerwünschten schwerwiegenden Ereignisse.	44
Tabelle 9	Übersicht der ausgewählten Studienzentren der NIVEAU-Studie.	45
Tabelle 10	Daten über die Patientenmerkmale bezogen auf die absolute Zahl der Meldungen. Aufgrund fehlender Normalverteilung sind Altersmerkmale als Median sowie 25%- und 75%-Quartil angegeben.	46
Tabelle 11	Die Tabelle zeigt die Verteilung der schwerwiegenden unerwünschten Ereignisse nach Systemorganklassen von CTCAE V5.0.	48
Tabelle 12	Schweregradeinteilung der schwerwiegenden unerwünschten Ereignisse aus den beiden Studienarmen durch lokale Studienärzte nach CTC-AE.	49
Tabelle 13	Kausalitätsbewertungen von unerwünschten schwerwiegenden Ereignissen aus dem experimentellen Arm der NIVEAU-Studie durch lokale Studienärzte, Beurteiler A und B.	51
Tabelle 14	Paarweiser Vergleich der Bewertungen anhand von dem Korrelationskoeffizienten von Spearman und Cohens Kappa für multivariable Entscheidungsform.	52
Tabelle 15	Die Kappa-Statistik sowie die prozentuale Übereinstimmung (%) zwischen alle 3 Beurteilerpaaren für die binäre Entscheidungsform (Sicher, wahrscheinlich, möglich – Ja, unwahrscheinlich, kein Zusammenhang – Nein).	53
Tabelle 16	Prozentuale Übereinstimmung zwischen Ratern. %EÜ = exakte prozentuale Übereinstimmung; %EU = exakte prozentuale Unübereinstimmung. %MV = minimale Variationen.	54
Tabelle 17	Die Kappa-Statistik sowie die prozentuale Übereinstimmung (%) zwischen den Beurteilern A und B bei der Evaluation der Erwartbarkeit.	55
Tabelle 18	Verteilung der ausgewählten Kausalitätskategorien nach Erwartbarkeit der Ereignisse eingestuft als solche durch den jeweiligen Beurteiler. Die Daten beziehen sich ausschließlich auf unerwünschte Ereignisse aus dem experimentellen Arm der Studie.	55
Tabelle 19	Kausalitätsbewertungsskala mit wahrscheinlichkeitsbasierten Kausalitätskategorien für die binäre Entscheidungsform abgeleitet von der WHO-UMC Skala.	63

# 1. Zusammenfassung und Abstract

## 1.1 Zusammenfassung

**Hintergrund:** Die Beurteilung des kausalen Zusammenhanges zwischen der Verabreichung des Studienmedikaments und dem Auftreten der schwerwiegenden unerwünschten Ereignisse besitzt eine hohe Sicherheitsrelevanz in klinischen Studien. Kausalitätsprüfung imitiert am häufigsten die Differenzialdiagnostik, wenn der Beurteiler alle zur Verfügung stehenden Informationen berücksichtigt und so den endgültigen Entschluss fasst. Dieses Vorgehen wird als Experteneinschätzung bezeichnet.

Die Daten über die Reliabilität der Experteneinschätzung stammen hauptsächlich aus dem Bereich außerhalb der klinischen Studien und sind oft kontrovers. Einige Besonderheiten machen aber die Kausalitätsprüfung in klinischen Studien unterschiedlich von den anderen Bereichen der Pharmakovigilanz. Das Ziel der vorliegenden Arbeit war die Prüfung der Übereinstimmung und Reproduzierbarkeit bei der Bewertung von schwerwiegenden unerwünschten Ereignissen in klinischen Studien.

Ein weiteres Ziel der Arbeit war die Untersuchung der Übereinstimmung bei der Evaluation der Erwartbarkeit der schwerwiegenden unerwünschten Ereignisse. Die Erkennung der schwerwiegenden Ereignisse, deren Art, Ausmaß oder Ergebnis von der Referenzsicherheitsinformation des Arzneimittels abweicht, ist essenziell für die kontinuierliche Nutzen-Risiko-Bewertung des Prüfpräparates. Die Evaluation von Erwartbarkeit war nach unserem Kenntnisstand noch nicht ein Gegenstand der Forschung.

**Methoden:** Hinsichtlich der Kausalität wurden insgesamt 177 schwerwiegende unerwünschte Ereignisse aus dem experimentellen Arm der NIVEAU-Studie durch zwei Beurteiler ausgewertet. Zur Evaluation der Übereinstimmung bei Erwartbarkeit wurden die Bewertungen von 272 Ereignissen aus beiden Studienarmen verglichen. Zur Bestimmung der Intrarater-Reliabilität wurden 117 Ereignisse durch den Beurteiler B in Abstand von 3 Monaten an 2 Zeitpunkten vorgenommen. Die Auswertung erfolgte anhand von Spearmans Korrelationskoeffizienten ( $\rho$ ) und ungewichtetem Cohens Kappa ( $\kappa$ ).

**Ergebnisse:** Der Beurteiler A wies mit lokalen Studienärzten ( $\kappa=0,227$ ,  $p<0,05$ ) sowie mit dem Beurteiler B ( $\kappa=0,347$ ,  $p<0,001$ ) nur eine geringe Übereinstimmung auf. Der Beurteiler B und lokale Studienärzte zeigten einen mittelmäßigen Übereinstimmungsgrad ( $\kappa=0,547$ ,  $p<0,001$ ) und eine moderate Korrelation der Bewertungen ( $\rho=0,600$ ,  $p<0,001$ ). Für beide Beurteiler vom Sponsor ( $\rho=0,456$ ,  $p<0,001$ ) sowie für den Beurteiler A und Studienärzte ( $\rho=0,332$ ,  $p<0,001$ ) war die Korrelation lediglich gering. Die Korrelation sowie die Übereinstimmung zwischen zwei

Bewertungen durch den Beurteiler B war sehr stark ( $\rho=0,912$ ,  $p<0.001$ ) bzw. sehr gut ( $\kappa=0,903$ ,  $p<0,001$ ).

Bei der Evaluation der Erwartbarkeit zeigten die Beurteiler A und B mit einem Kappa-Wert von 0,691 ( $p<0.001$ ) eine gute Interrater-Reliabilität. Der Übereinstimmungsgrad zwischen zwei Bewertungen durch den Beurteiler B war mit 0,680 ( $p<0,001$ ) ebenfalls gut.

**Schlussfolgerung:** Wir demonstrierten eine geringe Übereinstimmung sowie eine geringe Korrelation der Bewertungen von schwerwiegenden unerwünschten Ereignissen in einer Phase-III-Studie zwischen zwei Beurteilern von Sponsor sowie zwischen einem Beurteiler und lokalen Studienärzten. Nur der Beurteiler B demonstrierte eine moderate Korrelation bzw. eine mittelmäßige Übereinstimmung mit lokalen Studienärzten. Eine niedrige Korrelation bzw. Übereinstimmung zwischen Studienärzten und Beurteilern vom Sponsor kann zum Teil durch die niedrige Berichtsqualität erklärt werden. Hier lässt sich ein klares Verbesserungspotenzial erkennen.

Für die Übereinstimmung bei der Evaluation der Erwartbarkeit gibt es leider keine Daten zum Vergleich. Eine gute Übereinstimmung war zwar zu erwarten, die Prozentzahl der Fälle, die unterschiedlich bewertet wurden, war aber trotzdem mit 12,9 % bedenkenswert. Auffallend ist auch, dass die Übereinstimmung zwischen 2 Beurteilern fast genauso hoch ist wie zwischen 2 Bewertungen durch denselben Beurteiler. Diese Ergebnisse zeigen, dass auch bei der Evaluation der Erwartbarkeit ein Raum für unterschiedliche Interpretationen besteht, was weiterer Analyse bedarf.

## 1.2 Abstract

**Background:** The assessment of causality between study drugs and serious adverse events is of a high importance for safety in clinical trials. Causality assessment most often mimics differential diagnosis when the rater takes all the available information into account to reach the final conclusion. This process is known as expert judgement.

Data on the reliability of expert judgment is derived mainly from outside the clinical trials and are often controversial. However, some features make causality assessment in clinical trials different from the other fields of pharmacovigilance. The aim of the present work was to investigate the agreement and reproducibility in the assessment of serious adverse events in clinical trials.

Another objective was to investigate the agreement in the evaluation of expectedness of serious adverse events. Recognition of serious adverse events, type, magnitude, or outcome of which deviates from the reference safety information is essential for the ongoing risk-benefit assessment of the investigated drug. To our knowledge, evaluation of expectedness has not yet been a research subject.

**Methods:** Two raters evaluated the causality of 177 serious adverse events from the experimental arm of the NIVEAU trial. To evaluate the agreement for expectedness, the assessments of 272 events from both arms were compared. To determine the intra-rater reliability, 117 events were assessed by the rater B at two different time points 3 months apart. Spearman's correlation coefficient ( $\rho$ ) and unweighted Cohen's kappa ( $\kappa$ ) were used for analysis.

**Results:** Rater A demonstrated a low agreement with local investigators ( $\kappa=0.227$ ,  $p<0.05$ ) as well as with rater B ( $\kappa=0.347$ ,  $p<0.001$ ). The agreement was moderate between the rater B and local investigators ( $\kappa=0.547$ ,  $p<0.001$ ). For rater B and local investigators, the rank correlation was moderate as well ( $\rho=0.600$ ,  $p<0.001$ ). The correlation was low between both raters of sponsor ( $\rho=0.456$ ,  $p<0.001$ ) and for rater A and local investigators ( $\rho=0.332$ ,  $p<0.001$ ). Rank correlation as well as agreement between two assessments made by rater B was very strong ( $\rho=0.912$ ,  $p<0.001$ ) and very good ( $\kappa=0.903$ ,  $p<0.001$ ), respectively.

When evaluating expectedness, raters A and B demonstrated a good inter-rater reliability with a kappa value of 0.691 ( $p<0.001$ ). The level of agreement between two assessments from rater B was good as well at 0.680 ( $p<0.001$ ).

**Conclusion:** We demonstrated a low agreement and a low correlation of causality assessments of serious adverse events between two raters of sponsor and between one rater and local investigators. Only rater B demonstrated a moderate correlation and moderate agreement with local investigators. Low correlation and agreement between study investigators and raters of sponsor may be explained in part by low quality of reports offering a potential for improvement.

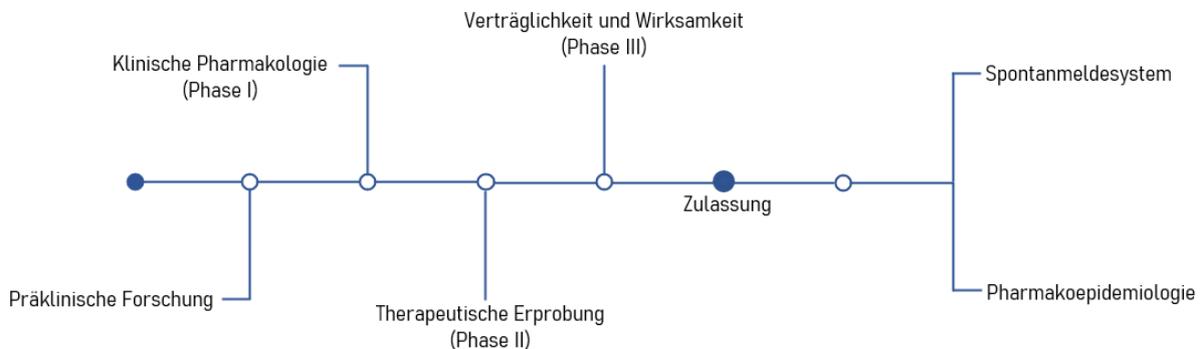
Unfortunately, for the agreement in the evaluation of expectedness, there is no data to compare. While a good agreement was to be expected, the percentage of cases assessed differently was still worth noting at 12.9%. Striking is, that the agreement between 2 raters is almost as high as between 2 assessments by the same rater. These results show that even in the evaluation of expectedness, there is a room for different interpretations, requiring further analysis.

## 2. Einleitung

### 2.1 Arzneimittelsicherheit in Klinischen Studien

Ein unerlässlicher Bestandteil der klinischen Arzneimittelprüfungen ist die aktuelle Nutzen-Risiko-Abschätzung, die auf dem Vergleich der bisher zur Verfügung stehenden sicherheitsrelevanten Informationen und dem zu erwartenden Nutzen basiert. Die Genehmigung einer klinischen Prüfung wird von der Ethikkommission und den Aufsichtsbehörden erst dann erteilt, wenn vom Sponsor alle nötigen Dokumente zur Gewährleistung der Prüfungssicherheit vorgelegt werden(28).

Fortlaufende und systematische Überwachung der unerwünschten Ereignisse eines Arzneimittels wird als Arzneimittelsicherheit (Pharmakovigilanz) bezeichnet(60). Die Erkennung von anwendungsbedingten Risiken von Arzneimitteln sowie allen Therapieformen ist eine Voraussetzung für ihre Marktzulassung. In dieser Hinsicht spielt die qualitätskonforme Durchführung der klinischen Studien eine entscheidende Rolle. Als Qualitätsstandard gilt dabei die Verordnung Nr. 536/2014 des Europäischen Parlaments und des Rates über klinische Prüfungen mit Humanarzneimitteln(26). Die Einhaltung der Prinzipien der EU-Verordnung ist ein Garant dafür, dass das Wohl und die Sicherheit der Prüfungsteilnehmer geschützt werden und die erhobenen Daten glaubwürdig sind(26).



**Abbildung 1** Fortlaufende und systematische Überwachung während des gesamten Lebenszyklus des Medikaments.

Das Sicherheitsmonitoring während Arzneimittelprüfungen unterliegt sowohl äußeren als auch inneren Regelungen. Die sicherheitsrelevanten Fragen werden streng durch das Studienprotokoll bestimmt, das als Schlüsselkomponente der wissenschaftlichen Studien betrachtet wird(60). Das Studienprotokoll bestimmt u. a. die Ein- und Ausschlusskriterien, die Art und den Zeitplan von relevanten klinischen oder Laboruntersuchungen sowie die Prinzipien der Dokumentierung und der Bearbeitung von unerwünschten Ereignissen (UE). Die Verantwortung für die Durchführung einer protokollgerechten Studie obliegt dem Sponsor(60).

Klinische Studien spielen eine bedeutende Rolle bei der Feststellung der Sicherheit der Arzneimittel. Ihre Überlegenheit gegenüber anderen Informationsquellen ist durch einige Faktoren bedingt:

- Für die Studien wird die Zielpopulation so gewählt, dass die Prüfungsergebnisse durch patientenbezogene Faktoren wie unerwünschte Medikamenteninteraktionen, Vorerkrankungen oder Risikofaktoren in sehr geringem Maße beeinflusst werden(60).
- Die Medikamentencompliance ist in klinischen Studien besser als bei pharmakoepidemiologischen Untersuchungen(60). Dies trifft besonders dann zu, wenn Medikamente in einer kontrollierten Umgebung wie z. B. im Krankenhaus oder in der Praxis verabreicht werden.
- Darüber hinaus ermöglicht die Besonderheit der klinischen Studien eine aktive Rückfrage und eine gewisse proaktive Erforschung seitens eines Studienarztes oder eines Sponsors, damit ein ursächlicher Zusammenhang festgestellt werden kann.
- Ein weiterer Vorteil besteht im Vorhandensein einer randomisierten Kontrollgruppe, die nach dem Vergleich der Studienarme und der statistischen Datenauswertung eine zuverlässige Kausalitätsanalyse ermöglicht(60).
- UEs werden im Rahmen der klinischen Studien nach strikt definierten Vorgaben erfasst und dokumentiert. Die hohe Qualität der erhobenen Daten kann die Evaluation des möglichen ursächlichen Zusammenhangs erleichtern(60).

Es muss aber bemerkt werden, dass klinische Studien gewisse intrinsische Limitationen aufweisen. Eine der Limitationen ist die Dauer der Studie, die sich häufig als ungenügend herausstellen kann, um unerwünschte Arzneimittelwirkungen (UAW) mit langer Latenzzeit zu erkennen(60).

Ferner verfügen klinische Studien aufgrund der unzureichenden Anzahl der Patienten oft nicht über ausreichende statistische Aussagekraft, um seltene unerwünschte Arzneimittelwirkungen zu erkennen. Laut der „Dreier-Regel“ („Rule of Three“) wäre z. B. für die Quantifizierung der UAWs, deren Frequenz bei 1:500 liegt, die Teilnahme von 1500 Patienten an der Studie notwendig(60). Außerdem wird die Studienpopulation normalerweise den konkreten Einschlusskriterien angepasst, weshalb sie manchmal nicht repräsentativ für andere Patientengruppen sein kann. Multimorbide Patienten oder Patienten aus Pflegeeinrichtungen werden z. B. oft von klinischen Studien ausgeschlossen(8).

Selbstverständlich ist die Sicherheit der Versuchsteilnehmer die allerwichtigste Aufgabe der klinischen Prüfungen. Um dies zu gewährleisten, müssen die innerhalb eines im Voraus festgelegten Zeitplans entstandenen unerwünschten Ereignisse kontinuierlich und rechtzeitig beschrieben werden(26). Dafür ist eine standardisierte Nomenklatur etabliert worden.

### 2.1.1 Unerwünschte Arzneimittelwirkung

Definition:

*„Als Nebenwirkung eines Medikaments, bezeichnet auch als unerwünschte Arzneimittelwirkungen (UAW) gilt jedes nachteilige Vorkommnis (inkl. abnormer Laborbefunde), welches beim Gebrauch eines Medikamentes entsteht und vermutlich mit dem Studienmedikament in Zusammenhang steht“*(24).

Mangelnde Wirksamkeit der Medikamente werden definitionsgemäß nicht als eine Nebenwirkung angesehen(11). Wichtig ist zu erwähnen, dass der Ausdruck „Reaktion auf ein Prüfpräparat“ in der Definition der Nebenwirkung bedeutet, dass ein kausaler Zusammenhang angenommen wird oder dieser zumindest nicht ausgeschlossen werden kann(24). Im Gegensatz dazu müssen bei unerwünschten Ereignissen (UE) Anhaltspunkte für den ursächlichen Zusammenhang nicht notwendigerweise vorliegen.

Der Begriff „Nebenwirkung“, etabliert im deutschen Sprachgebrauch, ist inhaltlich nicht mit „unerwünschter Arzneimittelwirkung“ identisch, da eine Nebenwirkung nicht zwangsläufig im eigentlichen Sinne des Wortes „unerwünscht“ ist. Dennoch werden unerwünschte Arzneimittelwirkung und Nebenwirkung oft als Synonym verwendet(11,12).

Je nachdem, ob eine Nebenwirkung dosisabhängig ist, erfolgt Ihre Einteilung in Typ-A- und in Typ-B-UAWs. Zu Typ-A-UAWs („Augmented Effects“) gehören die Ereignisse, die durch pharmakologisches Wirkspektrum der Medikamente bedingt sind(60). Dementsprechend sind sie auch dosisabhängig und reproduzierbar. Zu solchen Ereignissen gehören z. B. die Entwicklung der Insulin-induzierten Hypoglykämie oder die Blutung während der Antikoagulationstherapie. Die Typ-A-UAWs umfassen ca. 80 % der unerwünschten Arzneimittelwirkungen und werden oft noch in den klinischen Studien bzw. vor Marktzulassung beschrieben(60).

Typ-B-UAWs („Bizarre Effects“) haben hingegen keinen Bezug zu dem erwünschten Effekt der Medikamente und können folglich nicht durch seine Wirkweise erklärt werden(60). Daher sind sie dosisunabhängig und nicht reproduzierbar. Zu Typ-B-Nebenwirkungen gehören z. B. Hypersensitivitätsreaktionen, unter denen Anaphylaxien oder medikamentengetriggerte Autoimmunerkrankungen. Ihr Stellenwert nimmt wegen permanenter Entwicklung der neuartigen Therapieformen immer mehr zu. Es ist zu erwähnen, dass für die Erfassung der Typ-B-Ereignisse wegen ihrer Seltenheit die klinischen Studien nur eine begrenzte Bedeutung haben. Diese Rolle übernehmen die nach der Marktzulassung funktionierenden Spontanmeldesysteme und pharmakoepidemiologischen Studien(60).

Neben Typ-A und Typ-B Nebenwirkungen können wir auch folgende UAW-Kategorien erwähnen, die für mnemonische Zwecke folgendermaßen eingeteilt sind(60):

- Typ-C-UAW („Chronic Effects“): die nach der chronischen Einnahme der Medikamente entstandenen Ereignisse.
- Typ-D-UAW („Delayed Effects“): Ereignisse entstanden nach einer langen Latenzzeit nach dem Medikamentenverbrauch bspw. Tumorentstehung durch karzinogene Wirkung.
- Typ-E-UAW („End of Treatment Effects“): Ereignisse, die nach der Unterbrechung eines dauerhaften Medikamentengebrauchs auftreten, z. B. Rebound-Effekte nach dem plötzlichen Absetzen der Betablockertherapie oder Entzugssyndrome.

Der Tod von Patienten dürfen an sich nicht als Nebenwirkung eines Arzneimittels berichtet werden(11). Der tödliche Verlauf der Nebenwirkung ist aber ein Kriterium zur Einstufung des Events als „schwerwiegend“(26).

### 2.1.2 Unerwünschtes Ereignis

Definition:

*„Ein unerwünschtes Ereignis ist jedes nachteilige Vorkommnis, das einer betroffenen Person widerfährt, der ein Prüfpräparat verabreicht wurde und das nicht notwendigerweise in ursächlichem Zusammenhang mit dieser Behandlung steht“(26).*

Während der klinischen Studie muss jedes einzelne unerwünschte Ereignis (UE) unabhängig von angenommenem Kausalzusammenhang beschrieben und gemeldet werden(26). Unter dem Begriff - unerwünschtes Ereignis - wird nur ein zeitlicher Bezug zu der Verabreichung des Prüfpräparats gemeint, die Kausalitätsbestimmung rückt absichtlich in den Hintergrund(12,25,26). Darauf weist auch der Begriff „Ereignis“ hin, der im Unterschied zur „Reaktion“ inhaltlich keinen möglichen ursächlichen Zusammenhang vermittelt. Somit ist jede UAW ein unerwünschtes Ereignis, aber ein unerwünschtes Ereignis kann nur eine UAW werden, wenn ein positiver kausaler Zusammenhang angenommen wird. Jedes unerwünschte Ereignis, unabhängig davon, ob es mit dem Studienmedikament in Zusammenhang steht, muss vom Prüfer gemeldet und vom Sponsor in der Pharmakovigilanzdatenbank dokumentiert werden(12,28).

Ein unerwünschtes Ereignis stellt in klinischen Studien das niederschwelligste Sicherheitsereignis dar (*siehe Abbildung 2*)(10). Seine weitere Überarbeitung und Zuordnung gehören zu den Verpflichtungen des lokalen Studienarztes und des Sponsors(12,26).

### 2.1.3 Schwerwiegendes unerwünschtes Ereignis

Definition:

*„Schwerwiegendes unerwünschtes Ereignis (SUE) ist jedes unerwünschte Ereignis, das tödlich oder lebensbedrohend ist, eine stationäre Behandlung oder deren Verlängerung erforderlich macht oder zu bleibender oder schwerwiegender Behinderung oder Invalidität führt oder eine kongenitale Anomalie oder einen Geburtsfehler zur Folge hat“*(26).

Das Auftreten eines unerwünschten Ereignisses kann vom Prüfer bei der geplanten Patientenaufnahme während der Arzt-Patienten-Kommunikation erkannt werden oder der Patient selbst kann ihn darüber informieren. Unmittelbar danach muss der Prüfer aufgrund der eigenen Evaluation nach der Verordnung (EU) Nr. 536/2014 des Europäischen Parlaments und des Rates vom 16. April 2014 eine Entscheidung darüber fassen, ob das Ereignis als schwerwiegend bewertet werden kann(12,26). Wenn es von einem schwerwiegenden unerwünschten Ereignis handelt und ein begründeter Zweifel daran besteht, dass es durch Prüfpräparat hervorgerufen wurde, liegt, wie oben erläutert, eine schwerwiegende unerwünschte Arzneimittelwirkung vor, die bei der Durchführung der klinischen Studie zu grundsätzlichen Veränderungen wie etwa Medikamentendosierung, Studienpopulation und Überwachungsläufe führen kann(28). Die Entscheidung darüber, ob das unerwünschte Ereignis schwerwiegend ist, zählt zu den kritischsten Momenten bei der Durchführung der klinischen Studien.

Es ist zu bemerken, dass der Begriff „schwerwiegend“ nicht den Schweregrad oder die Ausprägung eines Ereignisses zeigt, sondern seine klinische Signifikanz(25). Belastungskopfschmerz kann z. B. für sich allein genommen sehr schwer sein, aber nicht als schwerwiegend eingestuft werden, denn sie können keine große Gefahr für einen Patienten darstellen. Somit sind schwerwiegende Ereignisse auch nicht mit „schweren“ (Grad 3) Ereignissen nach der Definition von Common Terminology Criteria of Adverse Events (CTC-AE) gleichzusetzen.

Ein unerwünschtes Ereignis wird als schwerwiegend betrachtet, wenn es unabhängig von der Dosierung:

- zum Tod eines Patienten führt
- lebensbedrohlich ist
- stationäre Behandlung oder Verlängerung einer bereits bestehenden Hospitalisierung erforderlich macht
- zu schwerwiegender Behinderung oder Invalidität führt
- eine kongenitale Anomalie oder einen Geburtsfehler zur Folge hat(25,26).

Laut ICH-Leitlinie wird unter dem Begriff „lebensbedrohlich“ ein Ereignis gemeint, bei dessen Auftreten das Leben des Patienten in Gefahr ist und nicht die Fälle, die hypothetisch den Tod verursachen könnten, wenn sie noch schwieriger gewesen wären(25). Als schwerwiegend müssen auch andere bedeutsame medizinische Ereignisse bewertet werden, die möglicherweise nicht unmittelbar lebensbedrohlich sind, keinen Tod verursachen oder keine Hospitalisation des Patienten benötigen, aber bei denen eine medizinische Intervention notwendig ist, damit man den oben genannten Ausgängen vorbeugen kann. Als Beispiel können allergische Bronchospasmen dienen, die zu Hause oder in der Notaufnahme behandelt werden müssen(25). Der Bewertung jedes einzelnen Falles muss eine medizinische und wissenschaftliche Beurteilung der Studienärzte und des Sponsors der Studie zugrunde liegen.

Sobald der Prüfer während einer Studie über ein schwerwiegendes unerwünschtes Ereignis erfährt, muss er innerhalb von 24 Stunden darüber den Sponsor in Kenntnis setzen(26). Ein häufiger Fehler, den der lokale Investigator begeht, ist, dass er auf die weitere Entwicklung des Krankheitsverlaufs wartet, damit nach der Erstmeldung keine s. g. Follow-up Meldung notwendig ist, was ein unzumutbarer Zeitverlust darstellt(28).

Nachdem der Prüfarzt das Auftreten des schwerwiegenden unerwünschten Ereignisses erkannt hat und diese als schwerwiegend eingestuft hat, müssen zwei wichtige Fragen beantwortet werden:

- a) War das Auftreten des Ereignisses bei der Anwendung dieses Arzneimittels zu erwarten?
- b) Wurde das Ereignis durch das Studienmedikament hervorgerufen?

#### 2.1.4 Bestimmung der Erwartbarkeit

Definition:

*„Eine unerwartete schwerwiegende Nebenwirkung ist eine schwerwiegende Nebenwirkung, die nach Art oder Schweregrad nicht mit der vorliegenden Information über das Prüfpräparat übereinstimmt“*(26).

Die Entscheidung darüber, ob das Ereignis zu erwarten war, basiert auf der s. g. Referenz-Sicherheitsinformation (RSI)(26). Diese ist entweder in Form einer s. g. Investigator's Brochure (IB) oder als Fachinformation dargestellt(25,28). Investigator's Brochure ist die Sammlung der aus bisher durchgeführten klinischen und präklinischen Prüfungen gewonnenen relevanten Daten über die Eigenschaften des Prüfpräparats(26). Wenn die

Prüfsubstanz bereits vermarktet ist und somit auch ihre pharmakologischen Eigenschaften längst bekannt sind, kann auf die Nutzung der umfangreichen Investigator's Brochure verzichtet werden. In diesem Fall kann die Fachinformation als Referenz genommen werden vorausgesetzt, dass sie alle aktuellen und detaillierten Informationen über das Studienmedikament enthält(10,15). CIOMS-Arbeitsgruppe VI empfiehlt, beim Medikamenteneinsatz für neue Indikationen, in neuer Dosierung oder in neuen Patientengruppen auch bei bereits zugelassenen Arzneimitteln die Investigator's Brochure zu bevorzugen(15).

Falls das Ereignis im entsprechenden Referenzdokument nicht bereits erfasst ist, wird es als unerwartet bewertet. Ausgangspunkt bei der Bewertung ist die Frage, ob das Ereignis bis zu diesem Zeitpunkt zu beobachten war und nicht das, was vom Arzneimittel aufgrund seiner pharmakologischen Eigenschaften oder von der Grunderkrankung selbst zu erwarten ist(25). Selbst dann, wenn das Ereignis sich vollständig durch die pharmakologische Wirkung des Studienmedikaments erklären lässt oder es eindeutig der Grunderkrankung zuzuordnen ist, kann das Ereignis als nicht erwartet bewertet werden. Der Begriff „Erwartbarkeit“ muss im regulatorischen Sinne und nicht im medizinischen Sinne verstanden werden(51).

Dabei ist wichtig, nicht nur die Art der Nebenwirkung zu berücksichtigen, sondern auch ihren Schweregrad. Wenn z. B. in der IB nur eine milde Pankreatitis beschrieben ist, aber aktuell ihre nekrotisierende Form vorliegt, hat man mit einem unerwarteten Ereignis zu tun(26). Als „nicht erwartet“ oder „nicht bekannt“ sollten auch die Ereignisse bewertet werden, die spezifischer sind als die in der Investigator's Brochure beschriebenen Ereignisse(10,26). Wenn z. B. in der Investigator's Brochure akutes Nierenversagen als UAW beschrieben ist, beim Patienten jedoch eine interstitielle Nephritis diagnostiziert wurde, muss sie als unbekannt gemeldet werden(25).

Die Vorgaben bezüglich des Inhaltes und der Form der Meldungen von lokalen Studienärzten an den Sponsor sind in der ICH-Leitlinie aufgeführt(25). Prüfärzte werden ermutigt, eine spezifische Diagnose zu stellen und diese an den Sponsor zu melden. Oft werden aber nur Krankheitssymptome oder Laborveränderungen berichtet, insbesondere dann, wenn noch keine endgültige Diagnose gestellt ist. Unstrukturierte Formulierungen lassen natürlich den gewissen Raum für unterschiedliche Interpretationen bei der Evaluation von Erwartbarkeit. CIOMS-Arbeitsgruppe VI empfiehlt, wenn möglich, zusammen mit Krankheitserscheinungen auch die Diagnosen zu liefern(15).

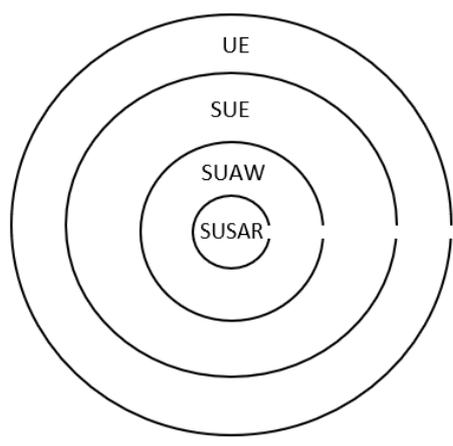
Falls, wie oben erläutert, ein schwerwiegendes unerwünschtes Ereignis auftritt, welches unerwartet ist und dabei noch zusätzlich der Zweifel besteht, ob es mit dem Prüfmedikament im Zusammenhang steht, muss es als Verdachtsfall einer unerwarteten, schwerwiegenden

Nebenwirkung (SUSAR) beschrieben und gemeldet werden(12,26). Bei einem SUSAR müssen daher folgende drei Voraussetzungen erfüllt werden:

- das Ereignis muss als schwerwiegend bewertet werden.
- das Ereignis muss im entsprechenden Referenzdokument nicht beschrieben und folglich unerwartet sein.
- es muss plausibel sein, dass das Ereignis im ursächlichen Zusammenhang mit dem Studienmedikament steht (bei einer Likert-Skala muss bspw. als Kausalitätskategorie „sicher“, „wahrscheinlich“ oder „möglich“ ausgewählt werden).

Wenn ein unerwünschtes Ereignis noch nicht als unerwünschte Arzneimittelwirkung des Medikaments beschrieben worden ist, kann das bedeuten, dass die Gesundheit der Prüfungsteilnehmer möglicherweise unter noch unbekanntem Risiko steht. Deswegen liegt es in der Verantwortung des Sponsors nach Bekanntwerden von SUSAR darüber die zuständige Ethikkommission, die zuständige Bundesoberbehörde (BfArM/PEI), Behörden anderer Mitgliedstaaten der Europäischen Union und anderer Vertragsstaaten des Abkommens über den Europäischen Wirtschaftsraum, in deren Hoheitsgebiet die klinische Prüfung durchgeführt wird, sowie die an der klinischen Prüfung beteiligten Prüfer in Kenntnis zu setzen(12,26). Die Frist dafür ist in den Richtlinien von GCP und EU-Parlament vorgeschrieben und umfasst maximal 15 Tage(12,26). Im Todesfall oder beim lebensbedrohlichen Ereignis beträgt sie jedoch maximal 7 Tage, die anderen relevanten Zusatzinformationen müssen in den nächsten 8 Tagen erteilt werden(12,26). Deshalb übernimmt der Sponsor auch die Verpflichtung, mindestens einmal im Jahr oder auf Anfrage der Ethikkommission und den zuständigen Bundesbehörden (BfArM/PEI) über die während der Studie entstandenen Verdachtsfälle unerwarteter schwerwiegender Ereignisse zu berichten(12,26).

Im Unterschied zur Kausalitätsbeurteilung - die vom Prüfer und Sponsor unabhängig voneinander durchgeführt wird - die Verantwortlichkeit für die Entscheidung darüber, ob das Ereignis erwartet war, trägt nur der Sponsor der Studie(10).



**Abbildung 2** Zusammenhänge zwischen verschiedenen Ereignissen in der Pharmakovigilanz. UE=unerwünschtes Ereignis, SUE=Schwerwiegendes unerwünschtes Ereignis, SUAW=schwerwiegende unerwünschte Arzneimittelwirkung, SUSAR=Verdachtsfall einer unerwarteten, schwerwiegenden Nebenwirkung.

### 2.1.5 Dokumentation und Meldung

Die Information über das mögliche unerwünschte Ereignis bekommt als Erster der lokale Studienarzt. Die erste Bewertung, Dokumentation und die Auskunftserteilung an den Sponsor sind seine direkten Verpflichtung(26).

Der entsprechende Fallberichtsbogen (Case Report Form, CRF) zur Dokumentation von unerwünschten Ereignissen wird vom Sponsor bereitgestellt (möglicherweise auch in elektronischer Form als eCRF). Der Fallberichtsbogen (der Prüfbogen) ist ein gedrucktes oder elektronisches Dokument, konzipiert zur Erfassung aller protokollrelevanten Daten, berichtet an den Sponsor(24). Dorr et al. demonstrierten, dass der Aufbau des Prüfbogens die Qualität der Berichterstattung erhöhen und Bewertungen der Ereignisse verbessern kann(20). Damit die UEs vollständig und ausführlich evaluiert werden können, müssen CRF bzw. eCRF folgende Bestandteile enthalten(28):

- die Bezeichnung des Ereignisses
- Beginn/Ende des Ereignisses
- Schweregrad
- Grund, warum das Ereignis als schwerwiegend betrachtet wird
- Stellungnahme über den möglichen ursächlichen Zusammenhang
- Angaben über die Verabreichung des Prüfpräparats sowie der Begleittherapie
- Alle weiteren Follow-up-Informationen (einschl. Outcome des Ereignisses).

Die Studienärzte sind verpflichtet, alle nicht schwerwiegenden UEs sachgemäß im CRF zu dokumentieren, die zeitnahe Informierung des Sponsors ist jedoch in der Regel nicht notwendig. Was aber schwerwiegende UEs betrifft, muss der Investigator in 24 Stunden oder

innerhalb eines Werktages nach ihrer Kenntnisnahme der Pharmakovigilanzabteilung des Sponsors eine Meldung erstatten(26).

Im Falle eines SUE wird die Meldung mithilfe eines speziellen Meldeformulars erstattet. Follow-up-Meldungen zu SUEs werden genauso wie Erstmeldung durch denselben Prüfbogen gemacht(28). Ein gesonderter Prüfbogen wird zusätzlich bei bestätigter Schwangerschaft gebraucht. Bei manchen klinischen Studien sind bestimmte Ereignisse als „Adverse Events of Special Interest“ (AESI) definiert. Diese werden auch in einem speziellen Meldebogen dokumentiert(28).

Nachdem der Sponsor die Auskunft von Studienarzt über ein SUE erhalten hat, muss der Sponsor es eigenständig und ausführlich dokumentieren und sie auf Anfrage an zuständige Behörden in Deutschland (BfArM/PEI) und anderen Mitgliedstaaten der Europäischen Union sowie in Vertragsstaaten des Abkommens über den Europäischen Wirtschaftsraum versenden, die an der klinischen Studie teilnehmen(26).

Die Nutzen-Risiko-Bewertung für die Gewährleistung der Patientensicherheit muss ununterbrochen erfolgen. Falls ein Zweifel an der Abweichung des Nutzen-Risiko-Verhältnisses während des Ablaufs der klinischen Studie entsteht, muss der Sponsor darüber die deutschen und internationalen Behörden innerhalb von 15 Tagen informieren. Zu den Sachverhalten, die dazu führen könnten, zählen(12,26):

- Einzelfallberichte über unerwünschte Arzneimittelwirkungen mit einem unerwarteten Ausgang
- klinisch relevante Erhöhung der Häufigkeit erwarteter schwerwiegender UAWs
- das Auftreten von SUSAR nachdem die Patienten die klinische Studie verlassen haben
- Ereignisse, die mit der Durchführung der klinischen Studie oder der Entwicklung des Studienmedikaments verbunden sind und die Gesundheit der Studienteilnehmer gefährden können.

## **2.2 Methoden der Kausalitätsbewertung**

Die Beschreibung möglicher Nebenwirkungen der Medikamente oder Medikamenteninteraktionen sowie ihre anschließende wissenschaftliche Analyse ist in klinischen Studien sowie im Postmarketing-Bereich der Kernbestandteil des Sicherheitsmonitorings. Ein kritischer Punkt dabei ist die Entscheidung, ob ein konkretes unerwünschtes Ereignis mit dem gegebenen Medikament in kausalem Zusammenhang steht. Diese Entscheidung erfordert ein hohes Maß an Verantwortung.

In den letzten Jahrzehnten wurden viele Methoden zur Standardisierung und Vereinfachung der Bewertung von Fallberichten erarbeitet(67). Die durch den Vergleich dieser Methoden gewonnenen Daten sind heterogen und manchmal sogar widersprüchlich. Einer der Gründe dafür ist, dass die für einen konkreten Zweck erarbeiteten Beurteilungssysteme nicht universell angewandt werden können(55). Die Medikamente können sich in pharmakologischen Eigenschaften stark voneinander unterscheiden und ihr Einsatz erfolgt auch unter unterschiedlichen Bedingungen. Die Validität der Kausalitätsbewertungsmethoden kann zudem nicht glaubhaft ermittelt werden(57). Dafür wäre ein zuverlässiger Standard notwendig, den es bei Kausalitätsprüfungen nicht gibt. Aus diesem Grund werden unter Umständen gewisse Ersatzstandards wie z. B. die konsensuellen Expertenbewertungen angewendet(47). Eine im internationalen Kontext approbierte Standardnomenklatur für die Kausalitätsbeurteilung existiert noch nicht(51). Es gibt auch keinen einheitlichen Standard für die Abstufung der Kausalität(25). Es werden oft Variationen der mehrstufigen Likert-Skala angewendet, darunter auch in den Formularen zur Meldung der UAWs an Bundesoberbehörden BfarM und PEI(9,58). Auch von WHO wird eine graduelle Antwortskala mit 6 Auswahlmöglichkeiten bevorzugt (*siehe Tabelle 1*)(67).

Die Methoden der Kausalitätsbewertung können in drei breite Kategorien unterteilt werden: Experteneinschätzung, Entscheidungsalgorithmen und Bayes'sche Analyse(51).

PEI	BfarM	WHO
	gesichert	certain
	wahrscheinlich	probable/likely
	möglich	possible
	unwahrscheinlich	unlikely
	unbeurteilt	conditional/unclassified
	nicht zu beurteilen	unassessable/unclassifiable

**Tabelle 1** Abstufung der Kausalitätskategorien durch PEI, BfarM und WHO(9,58,67).

### 2.2.1 Experteneinschätzung

Die Experteneinschätzung (engl. Global Introspection) ist eine individuelle wissenschaftliche Analyse der unerwünschten Ereignisse, die auf dem Wissen und Erfahrung eines Experten in einem konkreten medizinischen Fachgebiet sowie auf der Interpretation der einzelnen Fälle basiert(47). Der Prüfer ist verpflichtet, alle ihm zur Verfügung stehenden Informationen sorgfältig zu bewerten und die Schlussfolgerungen zu ziehen, was in der klinischen Praxis

dem Prozess der Diagnosestellung ähnelt. Da Schlussfolgerungen, die infolge der klinischen Beurteilung gezogen werden, keinen definierten Größenbezug haben können, wird das endgültige Urteil semiquantitativ gebildet.

Die theoretische Grundlage zur Kausalitätsbeurteilung zwischen zwei Ereignissen bildeten die im Jahr 1965 von Sir Austin Bradford-Hill in seiner Arbeit über Umwelteinflüsse auf die Gesundheit veröffentlichten Kriterien(35). Diese Kriterien umfassten: Assoziationsgrad, Konsistenz d. h. die Reproduzierbarkeit des Ereignisses, Spezifität, zeitlicher Zusammenhang, biologischer Gradient (Dosis-Wirkungs-Kurve), biologische Plausibilität, Kohärenz zwischen den vorhandenen Beweisen und aktueller Erscheinung, experimenteller Beweis und Analogie mit bereits bekannten anderen Nebenwirkungen. Die in den folgenden Jahrzehnten erarbeiteten unterschiedlichen Methoden der Kausalitätsbewertung basieren größtenteils auf den von Sir Austin Bradford-Hill vorgeschlagenen Kriterien, die oftmals verfeinert wurden(2).

Obwohl die Experteneinschätzung nicht einem bestimmten Standard folgt, sollten Beurteiler im Entscheidungsprozess die Kriterien berücksichtigen, die eine „intrinsische“ Evidenz für ihre Schlussfolgerungen bilden sollen. Die Kriterien, die im Entscheidungsprozess berücksichtigt werden sollten, können wir folgenderweise gruppieren(60):

- Pathophysiologische Plausibilität: manchmal ermöglicht der Wirkmechanismus des Studienmedikaments die Erklärung der Genese von UE. Im Falle von Immuncheckpoint-Inhibitoren wäre z. B. wegen des Risikos der Immuntoleranzstörung das Auftreten von Autoimmunerkrankungen richtungsweisend.
- Spezifische diagnostische Evidenz: Leider sind spezifische diagnostische Mittel meistens abwesend, ihre Existenz könnte allerdings einen zuverlässigen Hinweis liefern.
- Zeitlicher Zusammenhang: Die Latenzzeit zwischen der Administration des Arzneimittels und der Entstehung von unerwünschtem Ereignis ist anamnestisch meistens evident.
- Dechallenge und Rechallenge: Der positive und negative Dechallenge bzw. Rechallenge drückt die Abschwächung oder das Verschwinden des Ereignisses nach dem Absetzen der Therapie und seine erneute Entstehung oder Exazerbation im Falle der Wiederexposition aus.
- Vorhandensein bekannter Risikofaktoren: Die Angaben über die Vorerkrankungen eines Patienten sind wichtig, sodass man sie und ihre Komplikationen von unerwünschten Arzneimittelwirkungen unterscheiden kann.
- Ausschluss anderer Ursachen: UAW ist grundsätzlich eine Ausschlussdiagnose. In dieser Hinsicht ist die ausführliche Anamnese ausschlaggebend.

Arimone *et al* zeigten, dass bei Experteneinschätzung die Kriterien der intrinsischen Evidenz in sehr unterschiedlichem Maße zur abschließenden Bewertung der Ereignisse beitragen. Der zeitliche Zusammenhang war bspw. am häufigsten informativ, aber die Angaben über Dechallenge und Rechallenge spielten eine deutlich geringere Rolle(5). Hier muss allerdings angemerkt werden, dass bei zyklischer Medikamentenapplikation z. B. im Rahmen von klinischen Studien jeder erneute Therapiezyklus von Beurteilern als Rechallenge betrachtet werden kann(56). Es lässt sich folglich vermuten, dass die Kriterien abhängig vom Einsatzgebiet unterschiedlichen Stellenwert haben.

Unter extrinsischer Evidenz wird auf der anderen Seite der Vergleich mit in der wissenschaftlichen Literatur beschriebenen ähnlichen Fällen oder die Verwendung der Datenbank der Spontanmeldesysteme zur Feststellung der Kausalität gemeint(60).

Die auf einer nicht standardisierten Experteneinschätzung basierte s. g. „schwedische Methode“ wurde von Wiholm erarbeitet(66). Er unterscheidet zwischen sieben Faktoren, die bei der Feststellung des kausalen Zusammenhangs von Prüfern in Betracht gezogen werden sollten:

1. Zeitlicher Zusammenhang zwischen der Verabreichung des Medikaments und der Entstehung des Ereignisses
2. Ist das Reaktionsmuster eine bereits bekannte pharmakologische Wirkung des Medikaments oder deren Weiterentwicklung?
3. Abhängigkeit der klinischen Manifestationen von der Dosis. Sind sie weniger ausgeprägt bei reduzierter Dosierung?
4. Führt die Reexposition mit dem Medikament in gleichen Dosen zum erneuten Auftreten der Erscheinung?
5. Kann das Ereignis durch den klinischen Zustand des Patienten oder durch eine Begleiterkrankung des Patienten erklärt werden?
6. Kann das Ereignis durch die Begleittherapie oder durch die vorherige medikamentöse Behandlung erklärt werden?
7. Qualität der Beschreibung der aufgetretenen Symptome und der vorhandenen relevanten Laboruntersuchungen.

Anschließend wird der vom Experten bewertete kausale Zusammenhang entweder als P klassifiziert, was sowohl „Wahrscheinlich„ (Probable) als auch „Möglich“ (Possible) einschließt, oder als N, was sowohl „Unwahrscheinlich“ (unlikely), als auch „Nicht beurteilbar“ (Unassessable) umfasst.

Da die Experteneinschätzung auf der subjektiven Meinung eines Experten oder einer Expertengruppe beruht, produziert sie eine niedrige Beurteilerübereinstimmung, bezeichnet

auch als eine Interrater- bzw. Intrarater-Reliabilität, erklärt unter anderem durch unterschiedliche Kompetenzen und Erfahrungen der Beurteiler(5,43,46,57). Die Übereinstimmung zwischen Experteneinschätzungen und verschiedenen algorithmischen Methoden ist sehr variabel und übersteigt laut einigen Studien manchmal kaum die Zufallswahrscheinlichkeit(53).

Trotz der Nachteile entscheidet man sich am häufigsten wegen der Einfachheit für die nicht standardisierte Experteneinschätzung, unter anderem auch im Rahmen der klinischen Studien(15,51). Dabei sind die Angaben über die Validität dieser Methode sehr unterschiedlich und hängen von den zur Verfügung stehenden Informationen und den Bedingungen ab, unter denen die Bewertung erfolgt(55). Es konnte außerdem gezeigt werden, dass die Interrater-Reliabilität bei einzelnen Kriterien sowie zwischen einzelnen Kausalitätskategorien ziemlich hoch sein kann(5). Die Experteneinschätzung ist zudem befreit von einigen Limitationen der standardisierten Methoden. Deswegen werden auch bspw. die Extremkategorien häufiger ausgewählt als bei algorithmischen Methoden(63). Interessanterweise dienen verschiedene Variationen der Experteneinschätzung (wie z. B. konsensuelle Expertenzustimmung) zur Prüfung der Validität der algorithmischen Methoden oft als „goldener Standard“(63).

### 2.2.2 WHO-UMC Skala

Das Kollaborationszentrum der WHO für die Pharmakovigilanz (WHO-UMC) hat infolge der Beratung mit Nationalzentren unter Berücksichtigung der klinisch-pharmakologischen Aspekten ein standardisiertes Causality Assessment System für die Beurteilung von unerwünschten Arzneimittelwirkungen im Postmarketingbereich erarbeitet (*Tabelle 2*)(67).

Da eine der wichtigsten Aufgaben der Pharmakovigilanz die Entdeckung der immer noch unbekanntes Nebenwirkungen ist, rücken die Angaben über die schon beschriebenen Nebenwirkungen des Medikaments im Beurteilungssystem von WHO in den Hintergrund. Die Skala von WHO-UMC basiert auf 5 Kriterien(67):

1. Zeitlicher Zusammenhang zwischen der Verabreichung des Medikaments und des unerwünschten Ereignisses
2. Vorhandensein einer alternativen Ursache oder der Begleitmedikation
3. Reaktion auf das Absetzen des Medikaments oder die Dosisreduktion (Dechallenge)
4. Reaktion auf die erneute Verabreichung des Medikaments oder Dosissteigerung (Rechallenge);

„Sicher“ ist die einzige Kategorie, die ein zusätzliches Kriterium erfordert. Das Ereignis muss als pharmakologisch oder phänomenologisch definiert werden, wie z. B. das „Grey-Syndrom“ nach Administration von Chloramphenicol bei Neugeborenen(67). In diesem Fall ist das Rechallenge nicht notwendig, wenn andere Parameter glaubwürdig genug sind.

Nach den Kriterien der WHO-UMC Skala werden folgende Kausalitätskategorien unterschieden:

1. sicher
2. wahrscheinlich
3. möglich
4. unwahrscheinlich
5. nicht beurteilbar;

Kausalitätskategorie	Beurteilungskriterien
Sicher (Certain)	<ul style="list-style-type: none"> <li>▪ Plausibler zeitlicher Zusammenhang mit Verabreichung des Studienmedikaments.</li> <li>▪ Das Ereignis kann nicht durch Krankheit oder Komedikation erklärt werden.</li> <li>▪ Reaktion auf Absetzen der Behandlung ist pharmakologisch und pathophysiologisch plausibel.</li> <li>▪ Das Ereignis ist pharmakologisch und phänomenologisch definiert, was bedeutet, dass es eine spezifische Krankheit oder Störung oder ein bekanntes pharmakologisches Phänomen darstellt.</li> <li>▪ Rechallenge ist zufriedenstellend, falls notwendig.</li> </ul>
Wahrscheinlich (Probable/Likely)	<ul style="list-style-type: none"> <li>▪ Zeitlicher Zusammenhang mit Verabreichung des Studienmedikaments ist angemessen.</li> <li>▪ Es ist unwahrscheinlich, dass das unerwünschte Ereignis in Zusammenhang mit einem anderen Medikament oder einer Krankheit steht.</li> <li>▪ Reaktion auf das Absetzen der Therapie ist klinisch angemessen.</li> <li>▪ Reexposition ist nicht notwendig.</li> </ul>
Möglich (Possible)	<ul style="list-style-type: none"> <li>▪ Ereignis oder Laborabweichung mit einem angemessenen zeitlichen Zusammenhang mit Medikamenteneinnahme.</li> <li>▪ Kann auch durch eine andere Krankheit oder ein Medikament erklärt werden.</li> <li>▪ Information über die Reaktion nach dem Absetzen des Medikaments ist nicht verfügbar oder unklar.</li> </ul>
Unwahrscheinlich (Unlikely)	<ul style="list-style-type: none"> <li>▪ Ereignis oder Laborabweichung, deren zeitlicher Zusammenhang mit der Einnahme des Medikaments unwahrscheinlich ist.</li> <li>▪ Eine andere Krankheit oder ein anderes Medikament bieten eine plausible Erklärung.</li> </ul>
Unklassifiziert (Unclassified)	<ul style="list-style-type: none"> <li>▪ Ereignis oder Laborabweichung.</li> <li>▪ Die Information zur Prüfung der Kausalität ist nicht ausreichend.</li> <li>▪ Weitere Informationen unter Bearbeitung.</li> </ul>

Nicht beurteilbar (Unassessable)	<ul style="list-style-type: none"> <li>▪ Ein Bericht über eine mögliche UAW.</li> <li>▪ Beurteilung ist nicht möglich, weil die Information nicht ausreichend oder widersprüchlich ist.</li> <li>▪ Vervollständigung oder Überprüfung der Information ist unmöglich.</li> </ul>
-------------------------------------	---

**Tabelle 2** Die Kausalitätsbewertungsskala von WHO-UMC(67).

Wenn man sich mit dem Causality Assessment System von WHO auseinandersetzt, wird klar, dass seine Grundlage von den Protokollen der intrinsischen Evidenz gebildet wird. Die Kausalitätsabstufungen zeichnen sich in der Graduierung dadurch aus, wie ausgeprägt verschiedene Kriterien sind. Beim Vergleich von „Sicher“ und „Wahrscheinlich“ fällt beispielsweise auf, dass die Evaluation vom zeitlichen Zusammenhang fast identisch ist. Wenn sie im Falle von „Sicher“ plausibel sein muss, wird sie bei der Kategorie „Wahrscheinlich“ als angemessen (engl. Reasonable) definiert, was ein neutralerer Begriff ist und alles umfasst, was nicht „unangemessen“ erscheint(67). Der zeitliche Zusammenhang ist plausibel, wenn er darauf hinweist, dass das Ereignis durch das Medikament ausgelöst wurde, also es gibt ein positives Argument, welches dafür spricht. Hingegen ist der kausale Zusammenhang „unwahrscheinlich“, wenn der zeitliche Zusammenhang unglaubwürdig ist, was gegen die Kausalität spricht, wobei andere Medikamente und Krankheiten eine plausible Erklärung bieten. Die gleiche Gradierung der Wahrscheinlichkeit betrifft auch das andere Kriterium „Vorhandensein der alternativen Ursachen“ oder „Begleitmedikation“:

*Kann nicht erklärt werden – ist unwahrscheinlich – kann erklärt werden – bietet plausible Erklärung.*

Die Reexposition unter klinischen Bedingungen ist aus verschiedenen Gründen schwierig und oft sogar unmöglich. Daher ist seine Bedeutung in der WHO-UMC Skala minimiert und für keine Kategorie unbedingt notwendig. Die Information über Dechallenge ist aber viel öfter vorhanden und dieses Kriterium ist für die Klassifizierung von „sicher“ oder „wahrscheinlich“ unentbehrlich. Falls die Information bezüglich De- und Rechallenge nicht verfügbar ist, stehen nur zwei Kategorien zur Auswahl: „möglich“ oder „unwahrscheinlich“. Ihre Unterscheidung hängt vom zeitlichen Zusammenhang und davon ab, inwieweit es möglich ist, dass das Ereignis eine alternative Ätiologie hat.

Anhand dieser feinen inhaltlichen Differenzen können angrenzende Kategorien voneinander abgetrennt werden, was eine grundlegende Schwierigkeit bei der Feststellung der Kausalität darstellt. Da die häufigsten Beurteilungskategorien bei Meldungen von UEs „möglich“ und „wahrscheinlich“ sind, ist es ratsam, zunächst eine von diesen beiden Optionen als Referenz zu nehmen. Weitere „Herabstufung“ bzw. „Heraufstufung“ muss nach der Prüfung der jeweiligen einzelnen Kriterien erfolgen(67).

Das Bewertungssystem von WHO-UMC ordnet jede Kausalitätskategorie zu den Kriterien nach ihrem Ausprägungsgrad zu. Dadurch werden Bewertungen strukturiert und gleichzeitig wird die Grundlage zur Objektivierung geschaffen. Ohne dies werden solche Begriffe wie „möglich“, „wahrscheinlich“ oder „unwahrscheinlich“ unterschiedlich wahrgenommen, besonders in verschiedenen Sprachen(15). Außerdem sind „wahrscheinlich“ und „unwahrscheinlich“ im Grunde genommen „möglich“, durch die genaue Definition der Begriffe vermindert sich das Risiko der Überlappung. Die gegebenen Kriterien sind allerdings nur eine Anleitung, das Endfazit ist letztendlich aufgrund der individuellen Einschätzung zu ziehen. Deshalb wird die von WHO-UMC vorgeschlagene Methode als eine der Formen der Experteneinschätzung mit einer niedrigen Inter-Rater-Übereinstimmung assoziiert(18,62).

Die Anwendung der WHO-UMC-Kriterien nimmt wesentlich weniger Zeit in Anspruch als der Einsatz der schwierigen algorithmischen Methoden(62). Außerdem wird es im Gegensatz zu imperativen Entscheidungsalgorithmen durch den flexiblen Aufbau möglich, die Mehrzahl an Variablen bei der Urteilsbildung zu berücksichtigen, die in einem konkreten Algorithmus nicht vorgesehen wären. Möglicherweise kann dadurch auch eine bessere Interrater-Reliabilität der Kriterien von WHO-UMC im Vergleich zu algorithmischen Methoden in manchen Studien erklärt werden(55). Darüber hinaus ermöglicht die Skala von WHO-UMC die Beurteilung der Berichtsqualität, während bei algorithmischen Methoden die Kausalitätszuordnung erfolgt, auch wenn die Informationen nicht ausreichend ist(57).

### 2.2.3 Algorithmische Methoden

Da die Entscheidungsfindung nur auf Fachkenntnis und Erfahrung von Experten basiert, kann mglw. die objektive und zuverlässige Bewertung der unerwünschten Ereignisse gefährdet werden. Algorithmen stellen ein attraktives klinisches Instrument dar, welches durch eine Schritt-für-Schritt-Anleitung die strukturierte und standardisierte Bewertung der unerwünschten Ereignisse ermöglicht.

Eine der weitgehend akzeptierten algorithmischen Methoden für die Kausalitätsbewertung ist der Naranjo-Algorithmus(57). Naranjo *et al.* haben vor etwa 30 Jahren einen 10-Punkte-Fragenkatalog vorgestellt, der durch die den bestimmten Antworten zugeteilten Punkte ermöglicht, die Wahrscheinlichkeit davon zu berechnen, ob ein unerwünschtes Ereignis mit einem Medikament in ursächlichem Zusammenhang steht. Naranjo-Algorithmus gilt als erster Versuch der Standardisierung der Kausalitätsbewertung, der die Beurteilung von unerwünschten Ereignissen reproduzierbar gemacht hat. Im Gegensatz zum Bewertungssystem von WHO-UMC unterscheidet er nur 4 Kausalitätskategorien:

1. Definitiv (Definite)
2. Wahrscheinlich (Probable)
3. Möglich (Possible)
4. Zweifelhaft (Doubtful).

Bei mehrfacher Arzneimittelexposition sollte die Bewertung der Kausalität unabhängig für alle Medikamente erfolgen und das Medikament mit den meisten Punkten als mögliche Ursache des Ereignisses gewählt werden(57).

Fragen	Ja	Nein	Unbekannt
1. Gibt es frühere schlüssige Berichte von dieser Reaktion?	+1	0	0
2. Ist unerwünschtes Ereignis nach der Verabreichung des Medikaments entstanden?	+2	-1	0
3. Hat das Absetzen des Medikaments oder die Verabreichung des spezifischen Antagonisten zur Verbesserung vom UE geführt?	+1	0	0
4. Ist UE auch nach dem zweiten Verabreichen des Medikaments wiederaufgetreten?	+2	-1	0
5. Gibt es andere alternativen Ursachen, die das Ereignis auslösen könnten?	-1	+2	0
6. Konnte die Reaktion auch durch Placebo-Gabe ausgelöst werden?	-1	+1	0
7. Konnte das Medikament im Blut oder in anderen Körperflüssigkeiten in toxischer Konzentration nachgewiesen werden?	+1	0	0
8. War die Reaktion stärker bei der Erhöhung der Medikamentendosis und schwächer bei der Dosisreduktion?	+1	0	0
9. Hatte der Patient eine vergleichbare Reaktion auf dieses oder anderes Medikament bei der früheren Exposition?	+1	0	0
10. War UE objektiv nachweisbar?	+1	0	0

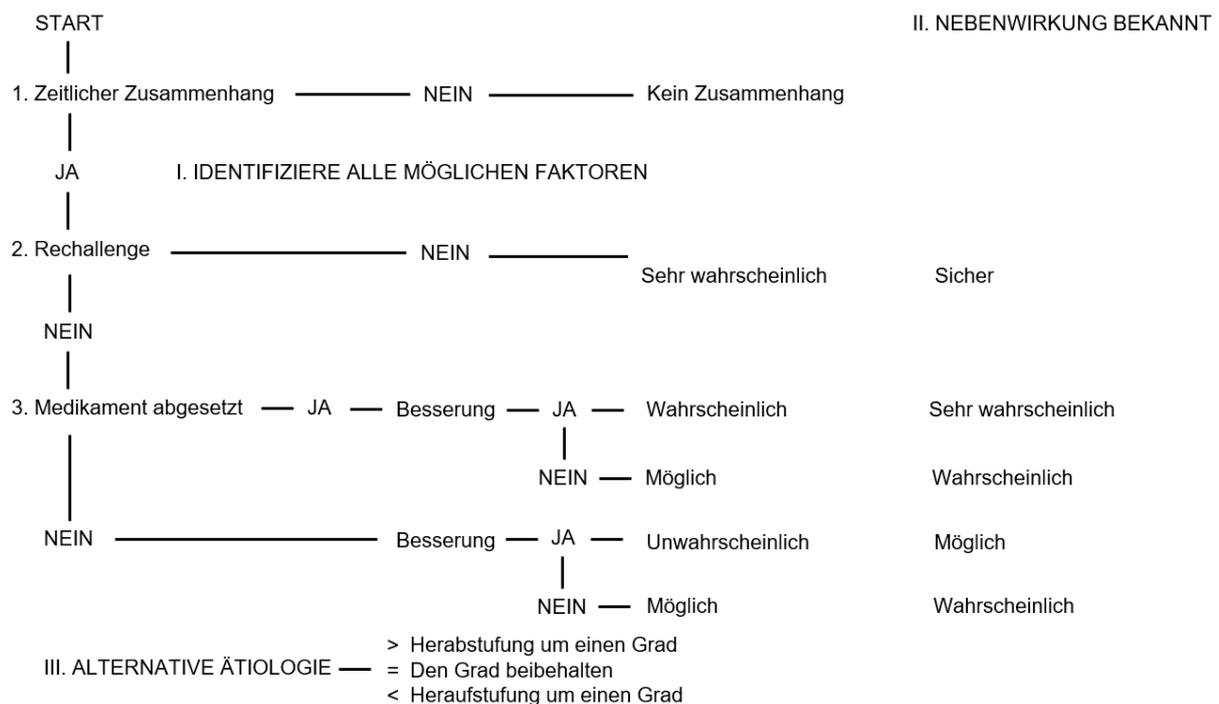
**Tabelle 3** Der Naranjo-Fragebogen und dessen Punkteverteilung(57). >9 = definitive UAW; 5-8 = wahrscheinliche UAW; 1-4 = mögliche UAW; 0 = zweifelhafte UAW.

Aufgrund von streng bestimmten Kriterien mit vorgegebenen Auswahlmöglichkeiten und entsprechender Punktebewertung ist die Interrater-Reliabilität sowie die Reproduzierbarkeit der algorithmischen Methoden verglichen mit Experteneinschätzung höher(57,62). In der Originalstudie schwankte die prozentuale Übereinstimmung bei Experteneinschätzung zwischen 41 % und 57 %, während sie nach dem Einsatz des Naranjo-Algorithmus bis 92 % stieg und die Intrarater-Übereinstimmung 97 % erreichte(57).

Für den Einsatz in klinischen Studien sind die algorithmischen Methoden nicht gut genug validiert(51). Von Ping-Hwa-Su und Ralph W. Stoll wurde 1993 zur Besserung der Reproduzierbarkeit und Reliabilität eine algorithmische Methode der Kausalitätsbewertung für lokale Studienärzte präsentiert(37). Der Algorithmus fungierte als Flussdiagramm und umfasste 3 Fragen:

1. Trat das Ereignis nach Medikamentengabe auf?
2. Reaktion auf De- und Rechallenge
3. Wurde die Behandlung aufgrund des Ereignisses abgebrochen?

Mithilfe zwei weiterer Kriterien - bekannte Assoziation mit dem Studienmedikament und alternative Ätiologie - wird eine genaue Zuordnung zu einer der 6 möglichen Kausalitätskategorien ermöglicht (siehe Abbildung 3). Die Methode von Ping-Hwa-Su und Ralph W. Stoll ist allerdings heute obsolet. In klinischen Prüfungen wird heutzutage hauptsächlich die Experteneinschätzung eingesetzt(15).



**Abbildung 3** Das Flussdiagramm zur Kausalitätsbewertung in klinischen Studien präsentiert von Ping-Hwa-Su und Ralph W. Stoll(37).

Algorithmische Bewertungssysteme führen mehr Variable in die Bewertung des Ereignisses ein. Das stellt für den Experten einerseits ein zusätzliches Entscheidungsfindungsinstrument dar, andererseits wird aber der Bewertungsprozess wegen der notwendigen Punktzuzuweisung und komplizierter Ablaufdiagramme zeitaufwendiger(62). Punkte werden zudem willkürlich zugewiesen, weswegen die Gewichtung der nominalen Fragen kritisch

hinterfragt werden sollte(57). Ferner lässt sich ein Teil dieser Kriterien oft nicht bewerten. Beispielsweise ist Rechallenge aus verschiedenen und oft auch aus ethischen Gründen, insbesondere bei schwerwiegender Arzneimittelwirkung, nicht vertretbar(5,19). Das Fehlen von Rechallenge, eines der wichtigsten Kriterien für den Naranjo Algorithmus, macht die Steigerung der Rangfolge der Kausalität auf maximal „wahrscheinlich“ möglich(19). Aufgrund von derartiger Abhängigkeit von bestimmten Kriterien kann manchmal, selbst wenn die Kausalität zu dem Medikament offensichtlich erscheint, die entsprechende Kategorie wegen der unzureichenden Punktzahl nicht gewählt werden(29). Im Endeffekt ist die weitaus häufigste Kausalitätskategorie bei Naranjo-Algorithmus „möglich“(29). Es muss weiterhin betont werden, dass auch Algorithmen nicht frei von subjektiven Einflüssen sind. Die Frage Nummer 5 in der Naranjo-Skala bezieht sich z. B. darauf, ob es alternative Ursachen gibt, die das Ereignis hervorrufen könnten, was von der Interpretation eines Experten abhängt. Zum Schluss, eine gute Übereinstimmung zwischen Experten bedeutet nicht notwendigerweise, dass die gewählte Option richtig ist. Dafür sind die algorithmischen ebenso wie alle Kausalitätsbewertungsmethoden wegen des fehlenden Goldstandards nicht gut genug validiert(51). Wegen dieser Mängel wird von manchen Autoren bei der Bewertung der Kausalität von schwerwiegenden Ereignissen vom Einsatz der algorithmischen Methoden abgeraten(6).

#### 2.2.4 Probabilistische Methoden

Probabilistische Methoden sind von dem Bayes-Theorem abgeleitet(51). Unter Bayes-Theorem versteht man die Aktualisierung der s. g. A-priori-Wahrscheinlichkeit - die man anhand von Hintergrundinformationen ermittelt hat - mit neuen verfügbaren Informationen. Daraus lässt sich die A-posteriori-Wahrscheinlichkeit des kausalen Zusammenhangs berechnen(39). Ein Nachteil der probabilistischen Methoden ist ihre niedrige Praktikabilität, da sie eine spezifisch kalkulierte Informationsdatenbank zur Berechnung der Wahrscheinlichkeit wie etwa UAW-Inzidenz voraussetzen. Deshalb ist ihre Bedeutung für klinische Studien niedrig.

### 2.3 Kausalitätsbewertung in klinischen Studien

Kausalitätsbewertung in klinischen Studien zeichnet sich durch einige Besonderheiten aus. Hinsichtlich der Arzneimittelsicherheit besteht der bedeutendste Unterschied darin, dass bei klinischen Studien nicht unerwünschte Arzneimittelwirkungen, sondern unerwünschte Ereignisse im Mittelpunkt stehen(26). Bei den Letzteren wird vom Berichtersteller die

Kausalität im Vorfeld nicht notwendigerweise angenommen(26). Im Postmarketing-Bereich wird dagegen von einer „impliziten Kausalität“ geredet, denn es wird kein Bericht erstattet, wenn der Zusammenhang mit dem Prüfpräparat nicht zumindest als möglich erscheint(25,51). Studienärzte sind ferner verpflichtet, die gesetzten Fristen einzuhalten, was in der Qualität des Berichts widerspiegeln kann, denn im Moment der Berichterstattung sind unerwünschte Ereignisse klinisch oft nicht eindeutig zuzuordnen(28). Bei Spontanmeldungen dagegen sind Ereignisse zum Meldezeitpunkt normalerweise bereits entfaltet. Außerdem trägt der meldende Prüfarzt im Unterschied zum Bereich außerhalb klinischer Studien Verantwortung, seine Stellungnahme über die Kausalität abzugeben(51). Patienten, eingeschlossen in klinische Prüfungen, leiden weiterhin öfter an verschiedenen Vorerkrankungen und sind selten symptomfrei noch vor der ersten Administration des Studienmedikaments. Cleeland et al. haben bspw. demonstriert, dass etwa ein Drittel der Studienpopulation bereits vor Einschluss in die Studie mindestens 3 mittelschwere oder schwere Krankheitssymptome aufwiesen(16). Es ist weiterhin zu bemerken, dass in klinischen Studien die Zuordnung der Kausalität signifikant durch studienspezifische Hintergrundinformation einschließlich der Patientencharakteristika beeinflusst werden kann. Beispielsweise wird bei Patienten mit kleinzelligem Lungenkarzinom pulmonale unerwünschte Ereignisse seltener dem Studienmedikament zugeschrieben(36).

Die Anzahl sowie die Art der verwendeten Kausalitätskategorien können zwischen klinischen Studien sehr unterschiedlich sein. Die Verwertbarkeit von multiplen Kausalitätskategorien in klinischen Studien wird häufig kritisch hinterfragt(38). Eaton et al konnten z. B. keine Korrelation der Anzahl von unerwünschten Ereignissen mit steigender Medikamentendosis für einzelne positive Kausalitätskategorien (Certain, Possible, Likely) feststellen(21). Die CIOMS VI Working Group empfiehlt auch eine binäre Entscheidungsform mit nur zwei Varianten: „Related“ oder „Not Related“(15). George et al. schlagen eine wahrscheinlichkeitsbasierte Formulierung der Kategorien vor mit „wahrscheinlich mit Studienmedikament assoziiert“ und „wahrscheinlich mit anderen Faktoren assoziiert“(30). Allerdings kann der Sponsor eigenständig entscheiden, wie die Kausalität eingestuft werden sollte, sodass nicht selten mehr als zwei benachbarte Kategorien eingesetzt werden (z. B. sicher, wahrscheinlich, möglich, nicht wahrscheinlich usw.)(51). Es ist nicht ratsam, in klinischen Studien Definitionen wie „unbekannt“ oder „kann nicht ausgeschlossen werden“ zu verwenden, denn dies kann zum „Overreporting“ und zur unnötigen Überlastung des Systems führen(15).

### 2.3.1 Kausalitätszuordnung durch Lokale Studienärzte

Im Rahmen der klinischen Studien erfolgt die erste medizinische Bewertung des unerwünschten Ereignisses durch den lokalen Prüfarzt beteiligter Zentren(26). Zu den Verpflichtungen des Prüfers gehört daher neben der Erkennung von UEs, ihrer klinischen Beurteilung und ggf. ihrer Einordnung als SUE auch eigene Stellungnahme in Bezug auf den möglichen kausalen Zusammenhang mit dem Prüfpräparat. Bei nicht schwerwiegenden UEs ist der Prüfarzt allerdings normalerweise nicht verpflichtet, seinen Standpunkt bezüglich der Kausalität zu dem Studienmedikament zu äußern. Die einzige Ausnahme können unerwünschte Ereignisse von substanzspezifischem Interesse (AESI) bilden(15).

In doppelblinden placebokontrollierten klinischen Studien muss der Experte immer davon ausgehen, dass der Patient zum experimentellen Arm gehört und dass er das Studienmedikament eingenommen hat(51).

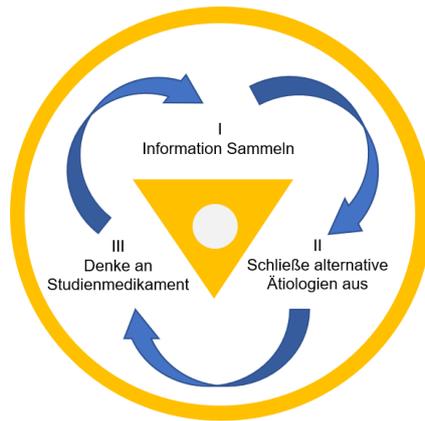
Die Kausalität wird vom lokalen Prüfarzt auf der Grundlage der Experteneinschätzung bestimmt(15). Die Bedeutung der Kausalitätsbewertung der unerwünschten Ereignisse in klinischen Studien jeglicher Phase kann nicht überschätzt werden. Anhand der Analyse der Kausalitätszuordnungen durch lokale Studienärzte wird die maximale tolerable Dosis ermittelt und die Bestimmung des Sicherheitsprofils des Studienmedikaments ermöglicht(40). Selbst die geringste Anzahl fehlerhafter Zuordnungen der Ereignisse zu dem Studienmedikament in Phase-I-Studien können in signifikanter Fehleinschätzung der maximal tolerierbaren Dosis (MTD) resultieren(40). Die inkorrekte Zuschreibung der unerwünschten Ereignisse zu dem Prüfarzneimittel kann zur Folge haben, dass es zu Dosismodifikationen kommt, das Indikationsspektrum geändert wird oder Medikamente im Extremfall vom Markt genommen werden. Werden Nebenwirkungen fälschlicherweise auf die Grunderkrankung oder Komedikation geschoben, kann die Patientensicherheit gefährdet werden.

Ping-Hwa Hsu und Ralph W. Stoll setzten sich mit der Problematik der Kausalitätsprüfung durch lokale Studienärzte auseinander, indem sie Bewertungen der unerwünschten Ereignisse aus 6 doppelblinden placebokontrollierten Studien analysierten(38). Es wurde gezeigt, dass Studienärzte die Kausalität bei Ereignissen aus der Experimentalgruppe häufiger positiv bewerteten, die Verteilung der Fälle, die als „Not Related“ bewertet wurden, war aber zwischen beiden Studienarmen vergleichbar. Eaton et al. untersuchten 38 Phase-I-Studien und demonstrierten eine signifikante Dosiskorrelation mit Toxizitäten für positive Kausalitätskategorien (sicher, wahrscheinlich und möglich), wobei für negative Kategorien (unwahrscheinlich, kein Zusammenhang) sie nicht zu beobachten war(21). Somit konnte gezeigt werden, dass die Kausalitätsprüfungen durch Studienärzte der Kausalitätserwartungen entsprachen.

Die Praxis der Kausalitätszuordnung durch lokale Ärzte in klinischen Studien wird allerdings oft kritisiert. Hillman et al. demonstrierten anhand der retrospektiven Analyse der doppelblinden randomisierten placebo-kontrollierten Phase-III-Studien, dass etwa die Hälfte der aus der Placebogruppe berichteten unerwünschten Ereignisse mindestens als „möglich“ klassifiziert wurden(48). Ferner wurden in bis zu 25 % der Fälle dieselben Ereignisse zu zwei unterschiedlichen Zeitpunkten anders bewertet. George et al. empfehlen zur Besserung der Kausalitätszuordnung durch lokale Studienärzte die Berücksichtigung von Kriterien der „intrinsic Evidenz“ bei Urteilsbildung und geben zusätzlich 4 Vorschläge(30). Diese können wir wie folgt zusammenfassen:

1. Nutzung einer binären Entscheidungsform
2. Verzicht auf Kausalitätszuordnung bei nicht schwerwiegenden Ereignissen in doppelblinden placebokontrollierten Studien
3. Bei Kombinationsstudien Kausalität für alle Medikamente zusammen einschätzen und nicht gesondert für das Studienmedikament
4. Einarbeitung eines effektiven Vorganges zur Entfernung der Ereignisse aus der RSI(30).

Zurzeit gibt es keine allgemein akzeptierte Methode für die Bewertung der schwerwiegenden unerwünschten Ereignisse in klinischen Studien. Dies wird unter anderem auch im Entwurf des Lebensmittelüberwachungs- und Arzneimittelbehörde der Vereinigten Staaten (FDA) zu guter Pharmakovigilanz-Praxis festgehalten(64). Trotz umfassender Kritik an der Experteneinschätzung bleibt sie auch in klinischen Prüfungen die weitestgehend beliebteste Herangehensweise. Mukherjee et al. führten 32 Interviews mit klinischen Onkologen, Hämatologen und Studienkoordinatoren mit langjähriger Erfahrung in Kausalitätsbewertung in klinischen Studien und zeigten, dass keiner der Befragten sich auf standardisierte Methoden verließen(56). Der typische Ablauf von Kausalitätsbewertungen imitierte größtenteils die Differenzialdiagnostik (*siehe Abbildung 4*). Von Ping-Hwa Hsu und Ralph W. Stoll konnten anhand der linearen Regressionsanalyse als die wichtigsten Kausalitätsfaktoren für lokale Studienärzte alternative Ätiologie und zeitlicher Zusammenhang identifiziert werden(38).



**Abbildung 4** Der übliche Vorgang der Kausalitätsbewertung der schwerwiegenden unerwünschten Ereignisse durch Studienärzte. Bild nach Mukherjee et al.(56).

2.3.2 Zweitbewertung der schwerwiegenden unerwünschten Ereignisse durch den Sponsor  
 Alle unerwünschten Ereignisse werden nach Kenntnisnahme und medizinischer Bewertung durch lokalen Studienarzt dem Sponsor zugeschickt. Als Nächstes erfolgt aufgrund der Information, die der Investigator geliefert hat, die unabhängige Zweitbewertung von UEs durch den Sponsor(26). Wenn das beschriebene Ereignis von dem Studienarzt oder vom Sponsor als schwerwiegend eingeschätzt und möglicherweise mit dem Studienmedikament in Verbindung gebracht wird, ist seine rechtzeitige Dokumentation und Analyse entscheidend für die kontinuierliche Risiko-Nutzen-Abwägung. Mängel in diesem Bereich könnten gravierende Folgen für Studienpopulation haben oder im massiven Anstieg der Studienkosten resultieren(56).

Aus Sicherheitsgründen darf der Sponsor keine Herunterstufung der Kausalität vornehmen oder die als schwerwiegend berichteten UEs als „nicht schwerwiegend“ umklassifizieren(15,26). Nichtsdestotrotz ist der Sponsor befugt, schwerwiegende unerwünschte Ereignisse erneuter Kausalitätsprüfung zu unterziehen und eigene Stellungnahme schriftlich zu verfassen(10,26).

Sponsors sind weiterhin befugt, eigenständig festzulegen, wie sie die Zweitbewertungen vornehmen. Eingesetzt werden standardisierte und nicht standardisierte Methoden, einschließlich bereits etablierter oder sogar selbst entwickelter einheimischer Algorithmen(15). Das gefällte Urteil ist von großer Bedeutung, denn nicht „erwartete“ SUEs können als SUSAR klassifiziert werden. Für die Einstufung des UE als SUSAR genügt nur eine positive Bewertung(25).

## 2.4 Rezidivtherapie des aggressiven B- und T-Zell-Lymphoms

### 2.4.1 Patienten geeignet für eine Hochdosistherapie mit autologer Blutstammzelltransplantation

Die Rezidivtherapie des aggressiven Lymphoms unterliegt zurzeit einer dynamischen Entwicklung. Dies ist begründet durch eine Vielzahl von neuen Substanzen und deren Prüfung in Phase-II und -III Studien. In diesem Kontext ist auch die NIVEAU-Studie zu sehen.

Bei jungen Patienten ohne therapielimitierende Komorbiditäten mit dem ersten Rezidiv des hochmalignen B- oder T-Zell-Lymphoms hat sich auf der Grundlage der Meilensteinstudie PARMA die Induktionstherapie gefolgt von einer Hochdosistherapie mit autologer Blutstammzelltransplantation als Behandlungsstandard etabliert(59). In der Rituximab-Ära wird als Induktion meistens die Immunchemotherapie mit (R)-DHAP (Rituximab mit Dexamethason, hochdosiertes Ara-C und Cisplatin) oder (R)-ICE (Rituximab mit Ifosfamid, Carboplatin, Etoposid) eingesetzt. Die beiden Regimen weisen etwa gleiche Therapieeffizienz auf, indem sie die Gesamtansprechrates von ca. 63 % erreichen(32). Dabei sind die Erfolgsaussichten nach allogener Blutstammzelltransplantation (ASZT) auf Patienten mit chemosensitiver Erkrankung begrenzt, die nach der Induktionstherapie eine partielle oder komplette Remission erreichen. Die Ansprechrates auf (R)-DHAP und (R)-ICE wird ihrerseits wesentlich durch den Zeitpunkt des Rezidivs bestimmt und ist fast zweimal so hoch bei Spätrezidiv als bei einem Rezidiv innerhalb von 12 Monaten (88 % bzw. 46 %)(32). Nach ASZT wird in 40 % der Patienten mit chemotherapiesensitiver Erkrankung eine langfristige Remission erzielt(59). Als Hochdosistherapie bietet sich ein BEAM-Protokoll, bestehend aus Carmustin, Etoposid, Ara-C und Melphalan(32).

Bei Patienten mit einem Rezidiv mit ungünstigen prognostischen Merkmalen in Form eines Progresses unter oder innerhalb eines Jahres nach Ende der Erstlinientherapie hat sich in zwei von drei Phase-III Studien (ZUMA-7, TRANSFORM, BELINDA) eine zelluläre Behandlung mit Chimeric Antigen Receptor (CAR) modifizierten T Zellen der Hochdosischemotherapie mit autologen Stammzelltransplantation als überlegen erwiesen. In der ZUMA-7 und der TRANSFORM-Studie war das Ereignis-freie Überleben als primäre Endpunkt signifikant besser, sodass eine Zulassung der Produkte in dieser Indikation erwartet wird(7,42,50).

#### 2.4.2 Patienten nicht geeignet für eine Hochdosistherapie mit autologer Blutstammzelltransplantation

Etwa 10 % der Patienten oberhalb des 60. Lebensjahres mit aggressiven Lymphomen sind primär refraktär und mehr als 25 % erleiden nach dem Ansprechen auf die Therapie ein Rezidiv(17). Insgesamt profitiert ca. die Hälfte der Patienten mit rezidivierender Erkrankung wegen des hohen Alters, der Therapieresistenz oder schwerer Komorbiditäten nicht von der Referenztherapie(33,61). Bei älteren Patienten ist die Anwendung von (R)-DHAP sowie der anderen dosisintensiven Therapieregime mit vergleichbaren Gesamtansprechraten wegen ihrer hohen Toxizität erschwert. In dieser Patientenpopulation gibt es eine einheitliche, durch die Phase-III-Studie bestätigte Therapiemodalität noch nicht(33,54).

#### 2.4.3 R-GemOx

Die nicht überlappenden Toxizitäten von (R)-GemOx sowie die Synergieeffekte einzelner Substanzen machen es zur attraktiven Therapieoption in Patienten mit rezidivierender oder refraktärer Erkrankung(27,33). Gemcitabin, ein Nukleosid-Analogon, verwandt mit Cytarabin und anderen Substanzen derselben Wirkstoffklasse, zeichnet sich durch relativ günstige pharmakokinetische Eigenschaften und ein breiteres Spektrum an Antitumoraktivität aus(65). Das Sicherheitsprofil von Gemcitabin ist dabei für die Mehrheit der Patienten durchaus akzeptabel(65). Gemcitabin kann auch als Monosubstanz bei Patienten mit schwer vorbehandelten aggressiven B- und T-Zell-Lymphomen eine objektive Ansprechrate von 20 %-71 % induzieren(14).

Im Vergleich zu Cisplatin besitzt das Drittgeneration-Cisplatinderivat Oxaliplatin nur minimale Oto- und Nephrotoxizität, was es zum attraktiven Kombinationspartner macht, insbesondere bei älteren und schwer vorbehandelten Patienten(14,54). Die Wirksamkeit von Oxaliplatin in rezidivierenden oder refraktären aggressiven B-Zell-Lymphomen konnte sowohl in platinbasierten Kombinationsregimen als auch als Einzelsubstanz nachgewiesen werden(49). Außerdem besitzt Oxaliplatin in Kombination mit Gemcitabin supraadditive Effekte, was noch vor 20 Jahren in leukämischen- und Kolonkarzinom-Zelllinien demonstriert wurde(27).

Die Wirksamkeit von R-GemOx in Patienten mit rezidivierenden oder refraktären B-Zell-Lymphomen nicht geeignet für eine Hochdosischemotherapie mit nachfolgender autologer Stammzelltransplantation wurde von einer multizentrischen Open-label-Phase-II-Studie untersucht. Nach 4 Zyklen erreichten 44 % der Patienten eine komplette und 17 % eine partielle Remission(54). Folglich betrug die Gesamtansprechrate in dieser Patientengruppe 61 %, was mit konventioneller platinbasierter Salvage-Therapie vergleichbar ist(54). Wegen

der hohen Ansprechrate von R-GemOx und insgesamt guter Verträglichkeit wird es als eine mögliche Plattform für die Kombination mit der zielgerichteten Tumorthherapie betrachtet. Dennoch ist die langfristige Prognose in Patienten mit einem rezidivierenden oder refraktären Non-Hodgkin-Lymphom ungünstig. Nach der R-GemOx Therapie beträgt das 5-Jahres-Gesamtüberleben nur 14 %, das mediane progressionsfreie Überleben liegt in mit Rituximab vorbehandelten Patienten lediglich bei 4 Monaten(54). Die Prognose ist noch schlechter beim Frührezidiv (<12 Monate), hier ist das mediane Gesamtüberleben fast viermal weniger als beim Spätrezidiv und beträgt nur 6 Monate(54).

### **3 Fragestellung und Zielsetzung**

Das Ziel der vorliegenden Arbeit war die Prüfung der Übereinstimmung bei der Evaluation der Kausalität der schwerwiegenden unerwünschten Ereignisse in klinischen Studien. Die bisherigen Daten über die Reliabilität der Experteneinschätzung basieren primär auf unerwünschten Arzneimittelwirkungen aus dem Bereich außerhalb der klinischen Studien. In klinischen Studien werden dagegen unerwünschte Ereignisse (UE) evaluiert, die sich durch einige Besonderheiten auszeichnen. Der Hauptunterschied besteht darin, dass bei UEs zu dem Zeitpunkt der Berichterstattung im Unterschied zu UAWs die Kausalität nicht notwendigerweise angenommen wird.

Die Kausalitätszuordnung in klinischen Studien wird oft kritisiert. Hillman et al. demonstrierten bspw., dass etwa die Hälfte der unerwünschten Ereignisse aus der Placebogruppe der doppelblinden randomisierten Phase-III-Studien mindestens als „möglich“ klassifiziert werden(48). Auf die Unzuverlässigkeit und Subjektivität der Kausalitätszuordnungen verweisen auch Enrico et al. aufgrund der systematischen Analyse der unerwünschten Ereignisse(23). Die Evaluation der Übereinstimmung ist die Voraussetzung für die Identifikation der Mängel bei der Bewertung der SUEs in klinischen Studien.

Ein weiteres Ziel der Arbeit war die Prüfung der Übereinstimmung zwischen Beurteilern bei der Evaluation von Erwartbarkeit. Die nicht erwarteten SUEs können je nach Kausalitätsbewertung als SUSAR klassifiziert werden, die in klinischen Studien eine hohe Sicherheitsrelevanz besitzen. Die Übereinstimmung bei der Evaluation der Erwartbarkeit war nach unserem besten Wissen noch nicht der Gegenstand der Forschung.

Die Fragestellungen der vorliegenden Arbeit können wir somit wie folgt zusammenfassen:

1. Wie hoch ist die Übereinstimmung bei der Kausalitätsbewertung von schwerwiegenden unerwünschten Ereignissen in klinischen Studien?

2. Was sind die Besonderheiten der unerwünschten Ereignisse im Hinblick auf Kausalitätsbewertungen?
3. Was sind die Ursachen der unterschiedlichen Kausalitätsbewertungen zwischen Studienarzt und Sponsor in klinischen Studien?
4. Wie hoch ist die Übereinstimmung bei der Evaluation der Erwartbarkeit?
5. Was könnte der unterschiedlichen Evaluation der Erwartbarkeit in klinischen Studien zugrunde liegen?

## 4 Material und Methoden

### 4.1 Synopse der NIVEAU-Studie

Studiennummer:	DSHNHL 2015-1
Kurztitel:	Niveau
Titel der Studie:	Verbesserung der Therapieergebnisse im ersten Rezidiv oder Progress eines aggressiven Non-Hodgkin Lymphoms bei älteren Patienten oder bei Patienten, die nicht für eine Hochdosischemotherapie infrage kommen, indem Nivolumab zu Gemcitabin, Oxaliplatin und Rituximab im Falle von B-Zell-Lymphom hinzugefügt wird.
Indikation:	Patienten mit aggressivem Non-Hodgkin-Lymphom im ersten Rezidiv oder Progress, die weder für eine Hochdosischemotherapie noch für eine autologe oder allogene Blutstammzelltransplantation geeignet sind.
Primäres Prüfziel:	Verbesserung des 1-Jahren-progressionsfreien Überlebens durch Nivolumab plus (R)-GemOX gefolgt von Nivolumab als Konsolidierung anstelle von alleinigem R-GemOx bei Patienten mit progressivem oder rezidiviertem aggressivem Non-Hodgkin-Lymphom, die sich weder für eine autologe noch allogene Blutstammzelltransplantation eignen.
Sekundäre Prüfziele:	<ul style="list-style-type: none"> <li>• Bestimmung, ob das Überleben durch Hinzugabe von Nivolumab zu Standard (R)-GemOx verbessert werden kann.</li> <li>• Bestimmung, ob das Outcome durch Hinzugabe von Nivolumab zu Standard (R)-GemOx verbessert werden kann.</li> <li>• Bestimmung der Toxizität und Protokolladhärenz von Standard (R)-GemOx mit oder ohne Nivolumab.</li> <li>• Evaluation der Lebensqualität der Patienten mit einem rezidivierten oder refraktärem Non-Hodgkin-Lymphom.</li> <li>• Analyse des Outcomes nach biologischen Parametern.</li> </ul>
Studiendesign, Statistik und Patientenzahl:	Internationale, multizentrische, randomisierte, unverblindete Therapieoptimierungsstudie mit vorausgehenden separaten Safety Run-In Phasen für B- und T-Zell-Lymphome.

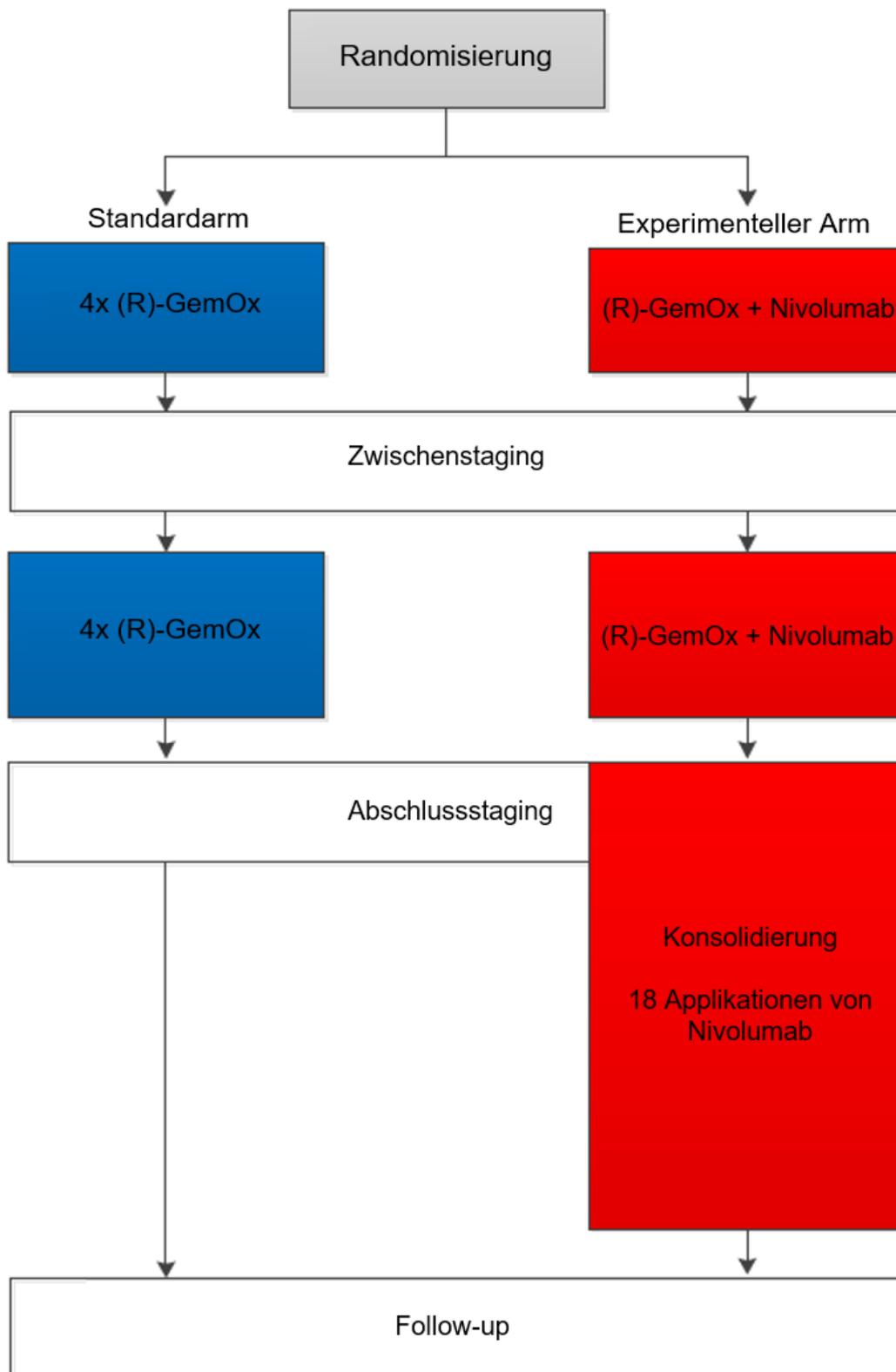
	<p>Das Ziel ist die Demonstration der Verbesserung des 1-Jahres-progressionsfreien Überlebens von 27 % auf 42 % (d. h. die Hazard Ratio von 0,66). Die zweiseitige Fragestellung wird mit einer Irrtumswahrscheinlichkeit von <math>\alpha=5\%</math> und einer Power 80 % beantwortet. Dazu wird die Analyse von 292 Patienten mit einem B-Zell-Lymphom (146 Patienten jeweils im experimentellen Arm und Standardarm) notwendig sein. Etwa 5 % der Patienten werden der Nachbeobachtung verloren gehen. Dementsprechend werden 310 Patienten mit einem B-Zell-Lymphom randomisiert. Gleichzeitig werden maximal 78 Patienten mit einem T-Zell-Lymphom in die Studie eingeschlossen.</p>
Studienpopulation:	<p>Patienten im ersten Rezidiv oder Progress eines aggressiven Non-Hodgkin Lymphoms, die sich weder für eine autologe noch allogene Stammzelltransplantation eignen, definiert durch Alter &gt;65 Jahre oder &gt;18 Jahre alt mit einem HCT-CI Score &gt;2 oder Patienten, die bereits eine autologe Stammzelltransplantation erhalten haben und nicht geeignet für eine allogene Stammzelltransplantation sind.</p>
Therapie:	<p>Immunochemotherapie besteht aus 8 Zyklen (R)- GemOx (Gemcitabin 1000 mg/m<sup>2</sup>, d1, Oxaliplatin 100 mg/m<sup>2</sup>, d1, im Falle eines B-Zell-Lymphoms: Rituximab 375 mg/m<sup>2</sup>, Wiederholung alle 2 Wochen)</p> <p><b>Standard Arm:</b> 8 Zyklen (R)-GemOx.</p> <p><b>Experimenteller Arm:</b> 8 Zyklen Nivolumab (240 mg) plus (R)-GemOx alle 2 Wochen gefolgt von zusätzlichen 9 Infusionen von Nivolumab (480 mg) in vierwöchigen Intervallen als Konsolidierung oder bis zur Progression oder inakzeptabler Toxizität.</p> <p>Während der Safety Run-in Phasen werden alle Patienten im experimentellen Arm behandelt.</p>
Primärer Endpunkt:	<p>1-Jahres-progressionsfreies Überleben.</p>

<p>Sekundäre Endpunkte:</p>	<ul style="list-style-type: none"> <li>• Rate kompletter Remissionen</li> <li>• Rate partieller Remissionen</li> <li>• Gesamtansprechrage</li> <li>• Dauer des Ansprechens</li> <li>• Progressionsrate</li> <li>• Rate therapieassoziierter Todesfälle</li> <li>• Rezidivrate</li> <li>• Ereignisfreies Überleben</li> <li>• Gesamtüberleben</li> <li>• Toxizität</li> <li>• Protokolladhärenz</li> <li>• Lebensqualität erfasst anhand von EQ-5D-5L</li> <li>• Therapieergebnis nach PD-L1 und PD-1 Expression, "cell of origin", 9p24.1 Alterationen</li> </ul>
<p>Analyse der Safety Run-in Phasen:</p>	<p>Die Durchführung der Sicherheitsanalysen ist nach dem Einschluss von 10 Patienten mit B-Zell-Lymphom und 5 Patienten mit T-Zell-Lymphom geplant. Für den Fall, dass die beabsichtigte Zahl von Patienten sowohl mit B-Zell-Lymphom als auch von Patienten mit T-Zell-Lymphom nicht rekrutiert wird, werden spätestens nach Einschluss entweder des 15. Patienten mit B-Zell-Lymphom oder des 7. Patienten mit T-Zell-Lymphom die Safety Analysen durchgeführt. Die Analyse wird durchgeführt, sobald der letzte Patient das Zwischenrestaging nach 4 Zyklen (R)-GemOx erhalten hat. Eine detaillierte Beschreibung der UEs/SUEs und therapie-assoziierten Todesfälle wird für B-Zell- und T-Zell-Lymphome getrennt erfolgen.</p> <p>Vor dem Einschluss weiterer Patienten in den randomisierten Therapieteil wird ein Substantial Amendment eingereicht.</p>
<p>Safety Analyse während der Randomisationsphase:</p>	<p>Die Analyse wird nach der Randomisierung von 30 Patienten in den experimentellen Arm erfolgen. Eine detaillierte Beschreibung der UEs/SUEs, Hämatotoxizität, Dosis von (R)-GEMOX und Nivolumab, Therapiedauer und therapie-assoziierten Todesfälle wird für beide Therapiearme (mit und ohne Nivolumab), ebenfalls für B- und T-Zell Patienten getrennt durchgeführt und dem Data Safety Monitoring Committee vorgelegt.</p>
<p>Hauptanalyse:</p>	<p>Alle Hauptauswertungen werden anhand des "intention-to-treat" Prinzips durchgeführt. Der Log-Rank Test wird in der ersten Analyse zum Vergleich des Progressions-freien Überlebens (PFS) der beiden Therapieformen</p>

	((R)- GemOx vs. Nivolumab plus (R)-GemOx gefolgt von einer Nivolumab Konsolidierung) verwendet. Um das PFS zu repräsentieren, werden Kaplan-Meier Kurven verwendet. Zusätzlich wird eine Projektion der 1-Jahres PFS Rate mit einem 95 % KI für die beiden Behandlungsstrategien erarbeitet. Zur Adjustierung der prognostischen Faktoren/Strata werden Cox Regressionsmodelle verwendet.
Zeitplan:	Insgesamt werden 310 Patienten mit B-Zell Lymphomen eingeschlossen. Wir erwarten eine langsamere Rekrutierung im ersten Jahr der Studie mit einer Rekrutierungsrate von 30 Patienten. In den folgenden Jahren gehen wir von der Rekrutierung von 70 Patienten pro Jahr aus, was in der Rekrutierungszeit von 5 Jahren resultiert, gefolgt von 2 Jahren Nachbeobachtung nach dem Einschluss des letzten Patienten. Mit einem erwarteten Rekrutierungsbeginn Q4/2017 wird die Rekrutierungszeit Q4/2022 und das Follow-up Q4/2024 abgeschlossen sein. Während des Rekrutierungszeitraums von 5 Jahren werden maximal 78 Patienten mit T-Zell Lymphomen eingeschlossen.
Sponsor:	Universität des Saarlandes, Saarbrücken, Deutschland.
Finanzielle Unterstützung:	Bristol-Myers Squibb.

**Tabelle 4** Synopse der NIVEAU-Studie(31).

## 4.2 Flussdiagramm der NIVEAU Studie



**Abbildung 5** Therapie der Patienten eingeschlossen in den experimentellen und Standardarm der NIVEAU-Studie (nach (34)).

### 4.3 Ein- und Ausschlusskriterien der NIVEAU Studie

Einschlusskriterien
<ul style="list-style-type: none"><li>• Patienten &gt; 65 Jahre</li></ul>
<ul style="list-style-type: none"><li>• Patienten &gt; 18 Jahre mit einem HCT-CI score &gt; 2</li></ul>
<ul style="list-style-type: none"><li>• Diagnose eines aggressiven Nicht-Hodgkin-Lymphoms basiert auf einer Exzisionsbiopsie von Lymphknoten oder der geeigneten Probe eines Extranodalbefalls bei der Initialdiagnose oder nach Rezidiv/Progression.</li></ul>
<ul style="list-style-type: none"><li>• Performance Status ECOG 0-2. Die Bestimmung von Performance Status erfolgt nach Prednisolon-Vorphase.</li></ul>
<ul style="list-style-type: none"><li>• Maximal eine bisherige anthracyklinhaltige Chemotherapie. Patienten mit einem B-Zell-Lymphom sollen Rituximab als Bestandteil des ersten Therapieregimes erhalten haben.</li></ul>
<ul style="list-style-type: none"><li>• Frauen im gebärfähigen Alter und sexuell aktive Männer müssen eine kontrazeptive Methode mit einer Fehlerrate &lt; 1 % Jahr anwenden.</li></ul>

**Tabelle 5** Zulassungskriterien zu der NIVEAU-Studie(34).

Ausschlusskriterien
<ul style="list-style-type: none"> <li>• Bereits initiierte Therapie nach dem ersten Rezidiv oder erster Progression</li> </ul>
<ul style="list-style-type: none"> <li>• Ernsthafte Nebenerkrankung oder Organdysfunktion (nicht Lymphom assoziiert): <ul style="list-style-type: none"> <li>- Angina pectoris CCS &gt;2, Herzinsuffizienz NYHA &gt;2</li> <li>- Gesamtbilirubin &gt;1,5 der oberen Normgrenze, Aspartataminotransferase oder Alaninaminotransferase &gt;3xULN</li> <li>- Kreatinin-Clearance &lt;30 ml/min</li> </ul> </li> </ul>
<ul style="list-style-type: none"> <li>• Leukozyten &lt;2500/μl, Segmentkernige Neutrophile &lt;2000/μl, Thrombozyten &lt;100.000/μl (dies gilt nicht, wenn die Zytopenie lymphombedingt ist).</li> </ul>
<ul style="list-style-type: none"> <li>• QTc-Intervall-Verlängerung &gt;450ms bzw. eine familiäre Vorgeschichte eines langen QT-Syndroms dokumentiert in einem Elektrokardiogramm. Dies gilt nicht bei einem kompletten Links- oder Rechtsschenkelblock.</li> </ul>
<ul style="list-style-type: none"> <li>• Patienten mit bestätigter oder vermuteter Autoimmunerkrankung. Zugelassen werden die Patienten mit Vitiligo, Diabetes mellitus Typ 1, residuale Hypothyreose aufgrund einer Autoimmunerkrankung, die lediglich eine Substitutionstherapie erfordert, Psoriasis, die keine Systembehandlung erfordert oder die Konditionen, bei denen kein Rezidiv in Abwesenheit eines externen Triggers erwartet wird.</li> </ul>
<ul style="list-style-type: none"> <li>• Keine Notwendigkeit einer immunsupprimierenden Therapie mit systemischen Kortikosteroiden (&gt; 10 mg/Tag Prednison-Äquivalent) für mindestens 2 Wochen vor der Administration des Studienmedikamentes.</li> </ul>
<ul style="list-style-type: none"> <li>• HIV-Infektion sowie eine chronisch aktive Hepatitis-B- oder C-Virus-Infektion.</li> </ul>
<ul style="list-style-type: none"> <li>• Patienten mit einem schweren Immundefizit.</li> </ul>
<ul style="list-style-type: none"> <li>• Stattgehabte Therapie mit Nivolumab, Gemcitabin oder Oxaliplatin.</li> </ul>
<ul style="list-style-type: none"> <li>• Patientin mit einem „gegenwärtig aktiven“ Zweitmalignom. Dies gilt nicht für nicht-melanotische Hautkrebsformen. Patienten mit dem Abschluss der Tumorthherapie vor mindestens 6 Monaten gelten nicht als diejenigen, die eine „gegenwärtig aktive“ Tumorerkrankung haben.</li> </ul>
<ul style="list-style-type: none"> <li>• Patienten mit einer primären oder sekundären ZNS-Beteiligung.</li> </ul>
<ul style="list-style-type: none"> <li>• Persistierende Neuropathie ab Grad 2 (CTC-AE v4.03).</li> </ul>
<ul style="list-style-type: none"> <li>• Schwangerschaft oder Stillzeit sowie die Frauen im gebärfähigen Alter.</li> </ul>
<ul style="list-style-type: none"> <li>• Persistierende Neuropathie Grad &gt;2 (CTC-AE v4.03) (dies gilt nicht bei Lymphom assoziierter Neuropathie).</li> </ul>

**Tabelle 6** Ausschlusskriterien der NIVEAU-Studie(34).

#### 4.4 Dokumentation und Meldung

Todesfälle, die durch die Progression des B-Zell-Lymphoms hervorgerufen sind, sowie die durch die Grunderkrankung ausgelösten Progressions- und Rezidivfälle stellen den primären Endpunkt der Studie dar. Deshalb werden solche Fälle nicht als SUE an Sponsor gemeldet. Im Gegensatz dazu werden Progressions- und Rezidivfälle des T-Zell-Lymphoms

dokumentiert und als SUE bewertet. Die Bewertungen der Kausalität und Erwartbarkeit dieser Meldungen wurden aus der Analyse ausgenommen.

Die Dokumentation von SUEs beginnt mit der ersten Verabreichung von Nivolumab und dauert 100 Tage nach der letzten Administration. Nach dem Ablauf dieser Frist kann jedoch die Dokumentation fortgesetzt werden, wenn der Zweifel besteht, dass das Ereignis durch das Studienmedikament ausgelöst wurde.

Die beteiligten Studienzentren sind laut dem Studienprotokoll verpflichtet, alle relevanten Informationen über den Krankheitsverlauf sowie die durchgeführte Behandlung gründlich und umfassend zu dokumentieren. Die Erfassung von schwerwiegenden SUEs erfolgt durch lokale Studienärzte auf einem elektronischen Prüfbogen (eCRF). Eintragungen in das eCRF erfolgen in Englisch. Im Studienprotokoll der NIVEAU Studie sind keine unerwünschten Ereignisse von besonderem Interesse (AESI) definiert. Der durch den Sponsor bereitgestellte Prüfbogen ist somit für alle SUEs einheitlich (*siehe Anhang*).

Lokale Studienärzte bewerteten den kausalen Zusammenhang des Studienmedikaments mit schwerwiegenden unerwünschten Ereignissen anhand der Likert-Skala präsentiert im eCRF. Zu den Verpflichtungen der Studienärzte gehörte weiterhin die Einteilung der Ereignisse in 5 Schweregrade nach CTC sowie die Angabe von Eventausgang und dem Grund für die Einstufung als schwerwiegend. Die Einstufung vom Schweregrad der Ereignisse erfolgte nach CTC-AE V4.03:

- Grad 1: Gering (engl. mild)
- Grad 2: mäßig (engl. moderate)
- Grad 3: schwer (engl. severe)
- Grad 4: lebensbedrohlich (engl. life-threatening or disabling)
- Grad 5: tödlich (engl. death related to the event)

Im eCRF der NIVEAU-Studie sind auf Grundlage der ICH-Leitlinie 6 Auswahloptionen für die Einstufung als schwerwiegend aufgelistet(25):

1. Fatal
2. Lebensbedrohlich
3. Notwendigkeit der stationären Aufnahme
4. Persistierende oder signifikante Behinderung
5. Angeborene Anomalie/Geburtsfehler
6. Andere signifikante oder medizinisch relevante Ereignisse (Siehe Studienprotokoll)

Der in der Niveau-Studie angewendete Prüfbogen enthält somit folgende Parameter, die bei der Bewertung von unerwünschten Ereignissen berücksichtigt werden sollen:

- Zugehörigkeit zu dem Studienarm
- Aktueller Therapiezyklus
- Angabe über den Zeitpunkt der ersten und letzten Verabreichung des Studienmedikaments
- Datum vom Auftreten und vom Ende des Ereignisses
- Schweregrad des Ereignisses nach CTC-AE
- Kausaler Zusammenhang mit dem Prüfpräparat
- Angabe darüber, ob das Ereignis schwerwiegend ist
- Ausgang des Ereignisses
- Relevante Vorerkrankungen
- Komedikation
- Relevante Laboruntersuchungen

Nachdem der Studienarzt über SUE erfahren hat, wird der ausgefüllte Meldebogen innerhalb von 24 Stunden per Fax an Leipziger Abteilung für Pharmakovigilanz des Zentrums für klinische Studien (ZKS) geschickt. Die Identifikation der Patienten und der Prüfzentren geschieht über die unikale Kennnummer.

Die Zweitbewertung erfolgt innerhalb von 2 Kalendertagen auf dem speziellen Beurteilungsbogen (Assessment Sheet) durch den Sponsor. Die Auswahloptionen, präsentiert im eCRF und Assessment Sheet, sind identisch und umfassen folgende Kausalitätskategorien:

1. sicher (engl. Certain)
2. wahrscheinlich (engl. Probable)
3. möglich (engl. Possible)
4. unwahrscheinlich (engl. Unlikely)
5. kein Zusammenhang (engl. Not Related)
6. nicht zutreffend (engl. Not Applicable)

Die 6. Auswahloption – „nicht zutreffend“ – wurde standardmäßig bei SUEs aus dem Standardarm gewählt. Die SUEs aus dem Standardarm wurden aus der Analyse der Kausalitätsbewertungen ausgenommen.

Auf dem Beurteilungsbogen müssen durch den Sponsor weiterhin folgende Fragen beantwortet werden:

1. War das Ereignis schwerwiegend?
2. Sicherheitsbelange, die die aktuelle Nutzen-Risiko-Bewertung ändern können?
3. Ausreichende Sicherheitsbelange für Änderungen in der Durchführung der Studie?

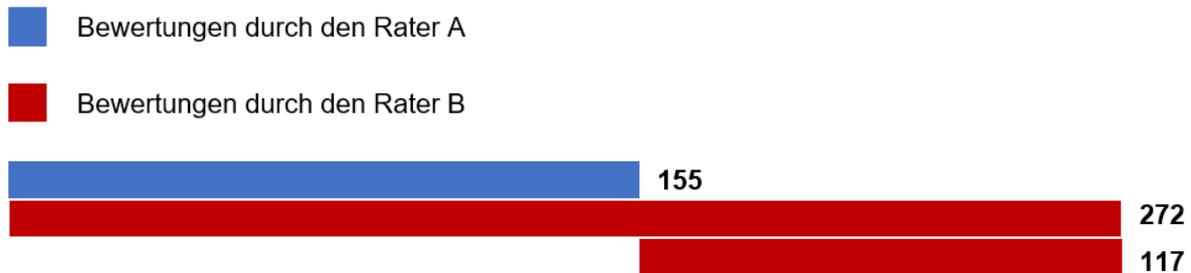
#### 4. War das Ereignis erwartet?

Eintragungen in den Beurteilungsbogen erfolgen ebenfalls in Englisch.

Alle Daten waren bereits durch den Sponsor der NIVEAU-Studie (Universität des Saarlandes) anonymisiert.

#### 4.5 Evaluation der Kausalität und Erwartbarkeit durch Rater

Zur Bestimmung der Interrater-Reliabilität wurden durch den Beurteiler A und B unabhängig voneinander 155 SUE Meldungen evaluiert, die im Zeitraum von 25.01.2018 bis 01.05.2020 aufgetreten waren. Der Beurteiler B bewertete hinsichtlich der Kausalität und Erwartbarkeit zusätzlich 117 SUEs, die im Folgezeitraum von 13.06.2020 bis 01.05.2021 auftraten (*Siehe Abbildung 6*). 83 SUEs davon waren aus dem experimentellen Arm.



**Abbildung 6** Die Abbildung zeigt die Verteilung der SUEs, die Beurteiler A und B hinsichtlich der Kausalität und Erwartbarkeit bewerteten. Zur Bestimmung der Intrarater-Reliabilität bewertete der Beurteiler B zusätzlich 117 SUEs zum zweiten Zeitpunkt im Abstand von mindestens 3 Monaten.

Aus den gesamten 272 SUEs waren 95 Ereignisse (34,9 %) aus dem Standardarm und 177 (65,1 %) Meldungen – aus dem experimentellen Arm. Die Entscheidung bezüglich der Kausalität in Verbindung mit dem Studienmedikament wurde nur für die Ereignisse aus dem experimentellen Arm getroffen, da im Standardarm das Studienmedikament nicht verabreicht worden ist. Die ausgewählte Kausalitätskategorie für SUEs aus dem Standardarm war daher standardmäßig „not related“. Dementsprechend wurden unerwünschte Ereignisse aus dem Standardarm zur Bestimmung der Interrater-Reliabilität bei Kausalitätsbewertung nicht berücksichtigt.

Der Beurteiler A und B bewerteten die Ereignisse ausschließlich anhand der nicht standardisierten Experteneinschätzung an der 5-stufigen Likert-Skala identisch wie im eCRF:

1. sicher (engl. Certain)
2. wahrscheinlich (engl. Probable)
3. möglich (engl. Possible)

4. unwahrscheinlich (engl. Unlikely)
5. kein Zusammenhang (engl. Not Related).

Kausalitätsbewertungen durch den Beurteiler A und B wurden verglichen miteinander sowie mit den Bewertungen von lokalen Studienärzten. Bei 3 Ereignissen fehlte die Angabe der Kausalität durch den lokalen Studienarzt. Somit wurden insgesamt 174 Kausalitätsbewertungen durch den lokalen Investigator analysiert.

Zur Evaluation der Intraraterreliabilität hat der Beurteiler B 117 unerwünschte Ereignisse zum zweiten Mal in der Phase 2 bewertet. Das Zeitintervall zwischen zwei Phasen betrug mindestens 3 Monate (*siehe Abbildung 6*).

Die Erwartbarkeit wurde bewertet unabhängig von der Zugehörigkeit zu dem Studienarm. Als Referenzsicherheitsinformation bei der Evaluation der Erwartbarkeit diente für beide Beurteiler ausschließlich die Fachinformation von Nivolumab.

#### **4.6 Statistische Methoden**

Die Beschreibung von kategorialen Variablen erfolgte als Absolutwerte und Prozent am Gesamtkollektiv. Kontinuierliche Variablen wurden zunächst mit dem Kolmogorov-Smirnov-Anpassungstest und Shapiro-Wilk-Test auf Normalverteilung geprüft. Daten, die nicht normalverteilt waren (Patientenalter, Zahl der verabreichten Zyklen im experimentellen Arm) wurden als Median sowie 25%- und 75%-Quartil angegeben.

Zur Bestimmung der Interrater-Reliabilität wurde Cohens Kappa ( $\kappa$ ) angewandt. Bei ordinalskalierten Daten erfolgte die Berechnung der Rangkorrelation zwischen den Bewertungen anhand von dem Korrelationskoeffizienten nach Spearman (auch Spearman-Rho oder  $\rho$ ).

Die Merkmale Erwartbarkeit und Kausalität wurden auf die stochastische Unabhängigkeit anhand eines Chi-Quadrat-Tests geprüft.

Alle p-Werte wurden zweiseitig getestet. Ergebnisse mit  $p < 0,05$  wurden als statistisch signifikant gewertet. Die entsprechenden p-Werte sind sowohl bei Cohens Kappa als auch bei Korrelationskoeffizienten nach Spearman mit angegeben. Alle statistischen Auswertungen erfolgten mit dem SPSS-Statistikprogramm (Version 28.0, IBM, Chicago, Illinois, USA).

Kappa-Wert	Grad der Übereinstimmung	Spearman Rho	Grad der Korrelation
<0	Niedriger als Zufall	0	Keine Korrelation
0,01-0,20	Schwach	0,1-0,29	Schwach
0,21-0,40	Gering	0,3-5,9	Gering
0,41-0,60	Mittelmäßig	0,6-7,9	Moderat
0,61-0,80	Gut	0,8-9,9	Sehr stark
0,81-0,99	Sehr gut	1	Perfekt

**Tabelle 7** Qualitative Bewertungen der Kappa-Werte sowie der Korrelationskoeffizienten nach Spearman im medizinischen Bereich(13).

## 5 Ergebnisse

### 5.1 Beteiligte Zentren

An der NIVEAU-Studie nehmen 75 Studienzentren aus 9 Ländern teil. Wir analysierten 272 SUE Meldungen aus 49 Studienzentren in 7 Ländern. Die große Mehrheit der schwerwiegenden Ereignisse stammte aus Deutschland (n=157, 57,7 %). Den zweiten Platz nahm Belgien mit 43 Meldungen (15,8 %) ein (*siehe Tabelle 8*). Das Studienzentrum BE 680 war für die meisten Meldungen (n=20) an den Sponsor verantwortlich. Die zweite Stelle teilten mit jeweils 19 Meldungen die Zentren DE 012 und DE 337. Ausgewählte Studienzentren sind in Tabelle 9 aufgeführt.

Land	Häufigkeit	Prozent
Deutschland	157	57,7
Belgien	43	15,8
Frankreich	42	15,4
Österreich	14	5,1
Niederlande	10	3,7
Portugal	5	1,8
Polen	1	0,4
Gesamt	272	100 %

**Tabelle 8** Herkunftsländer der unerwünschten schwerwiegenden Ereignisse.

Land-ID	Zentrum-ID	Häufigkeit	Prozent
BE	680	20	7,4
DE	12	19	7,0
DE	337	19	7,0
DE	73	15	5,5
DE	306	14	5,1
DE	52	12	4,4
DE	43	12	4,4
DE	113	9	3,3
BE	759	8	2,9
Rest		144	53
Summe		272	100

**Tabelle 9** Übersicht der ausgewählten Studienzentren der NIVEAU-Studie.

## 5.2 Beschreibung des Patientenkollektivs

Das mediane Lebensalter der Patienten zum Zeitpunkt der Berichterstattung lag bei 77 Jahren. Der jüngste Patient war 53 Jahre, der älteste 87 Jahre alt. In 2 Meldungen fehlte die Angabe über das Alter. 177 SUE Meldungen (65,1 %) kamen von den Patienten aus dem experimentellen Arm der Studie und 95 SUEs (34,9 %) von den Patienten, die in den Standardarm randomisiert waren. 175 Ereignisse (64,3 %) traten bei den Patienten auf, die zum Zeitpunkt der Meldung unter einem B-Zell-Lymphom litten. 92 SUE Meldungen (33,8 %) betrafen die Patienten erkrankt an einem T-Zell-Lymphom. Bei 5 Meldungen (1,8 %) fehlte die Angabe zum Lymphomtyp. In 134 Meldungen waren Patienten Frauen (49,3 %) und in 136 Meldungen (50,0 %) Männer. In 2 SUE Meldungen fehlte die entsprechende Angabe.

Es ist zu bemerken, dass die oben aufgeführten Daten sich auf die schwerwiegenden unerwünschten Ereignisse beziehen, die in unserer Arbeit eingeschlossen wurden und nicht auf die absolute Zahl der Patienten, eingeschlossen in der NIVEAU-Studie. Somit sind die Daten nicht oder nur teilweise für die gesamte Studienpopulation repräsentativ. Alle statistischen Auswertungen wurden zudem unabhängig von dem Studiensponsor ausgeführt.

Die für die statistische Analyse ausgewählten 272 SUEs stammten von insgesamt 118 Patienten. Entsprechend kamen im Durchschnitt 2,3 Meldungen von demselben Patienten/derselben Patientin. 68 Patienten erlitten mehr als ein unerwünschtes Ereignis. Die meisten unerwünschten Ereignisse erlitten Patienten NIV-0080 und NIV-0213 (n=8). Aus 118 Patienten waren 61 (51,7 %) Frauen und 57 (48,3 %) Männer. Die Verteilung der Patienten in 2 Studienarmen war identisch mit den Daten bezogen auf die gesamte Anzahl der SUE-Meldungen (65,4 %/34,6 %). Die Tabelle 10 zeigt im Überblick die Patientenmerkmale bezogen auf die gesamte Anzahl der ausgewerteten SUE-Meldungen zum Zeitpunkt der Berichterstattung.

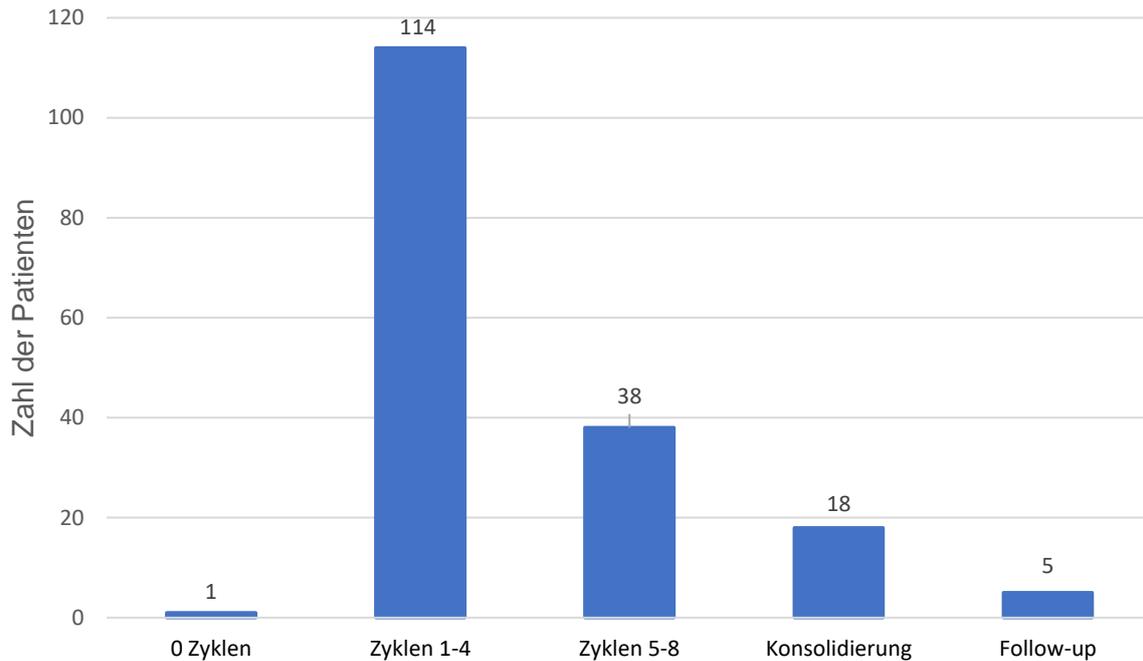
<b>Merkmal</b>	<b>Wert</b>	<b>Prozent</b>
Alter	77 [72;79]	
Männliches Geschlecht	136	50,0
Weibliches Geschlecht	134	49,3
Experimenteller Arm	177	65,4
Standardarm	94	34,6
B-Zell-Lymphom	175	64,3
T-Zell-Lymphom	92	33,8

**Tabelle 10** Daten über die Patientenmerkmale bezogen auf die absolute Zahl der Meldungen. Aufgrund fehlender Normalverteilung sind Altersmerkmale als Median sowie 25 %- und 75 %-Quartil angegeben.

### **5.3 Charakterisierung der Schwerwiegenden unerwünschten Ereignisse**

Aus dem gesamten 272 SUEs waren 95 (34,9 %) aus dem Standardarm und 177 (65,1 %) - aus dem experimentellen Arm.

Zum Zeitpunkt der SUE Meldung wurden im Median bereits 3 Zyklen [1;7] verabreicht. Die Anzahl der Therapiezyklen variierte zwischen 0 und 23. 5 SUEs (2,8 %) traten bei Patienten auf, die sich im Follow-up befanden. Ein SUE trat nach dem Einschluss in die Studie, aber noch vor der Verabreichung des Studienmedikaments auf. Bei einer SUE Meldung fehlte die Angabe bezüglich des Therapiezyklus (*siehe Abbildung 7*). Es ist zu bemerken, dass diese Daten sich auf die 272 Ereignisse beziehen, die wir zur statistischen Auswertung ausgewählt haben. Somit sind sie nicht für die gesamte Studie repräsentativ. Die Datenauswertung für diese Arbeit erfolgte unabhängig von dem Sponsor der Studie.



**Abbildung 7** Das Balkendiagramm zeigt, in welcher Phase bei den Patienten aus dem experimentellen Arm der Studie die SUEs auftraten.

Wir kategorisierten schwerwiegende unerwünschte Ereignisse nach den Common Terminology Criteria for Adverse Events (CTCAE). Die gesamten 272 schwerwiegenden unerwünschten Ereignisse wurden in 20 Organklassensysteme eingruppiert. Infektionen und Infestationen traten mit 26,5 % (n=72) am häufigsten auf. Unter Infektionen teilten die ersten 2 Plätze jeweils mit 22,2 % (n=16) Lungeninfektionen und Sepsis. In 19,4 % der Fälle (n=14) konnte kein Infektfokus festgestellt werden.

Den zweiten Platz nahmen mit 19,5 % (n=53) allgemeine Störungen und anwendungsbedingte Umstände ein. Darunter war das Fieber mit großem Abstand (n=30, 56,6 %) das häufigste unerwünschte Ereignis. Den zweiten Platz nahm mit 22,6 % der reduzierte Allgemeinzustand (n=12) ein. Das dritthäufigste Organklassensystem mit 10,7 % (n=29) war der Gastrointestinaltrakt.

<b>Systemorganklassen</b>	<b>n (%)</b>
Infektionen und Infestationen	72 (26,5)
Allgemeine Störungen und anwendungsbedingte Umstände	53 (19,5)
Gastrointestinale Störungen	29 (10,7)
Untersuchungen	19 (7,0)
Atemorgan-, Thorax- und Mediastinalstörungen	16 (5,9)
Kardiale Störungen	13 (4,8)
Vaskuläre Störungen	10 (3,7)
Stoffwechsel- und Ernährungsstörungen	9 (3,3)
Verletzungen, Vergiftungen und prozedurale Komplikationen	8 (2,9)
Nervensystemstörungen	7 (2,6)
Immunsystemstörungen	7 (2,6)
Blut- und Lymphsystemstörungen	6 (2,2)
Nieren- und Harnwegsstörungen	6 (2,2)
Hepatobiliäre Störungen	5 (1,8)
Muskel-, Skelett- und Bindegewebestörungen	5 (1,8)
Ohr- und Labyrinthstörungen	2 (0,7)
Neoplasien, gutartig, bösartig und unspezifiziert (inklusive Zysten und Polypen)	2 (0,7)
Endokrine Störungen	1 (0,4)
Haut- und Subkutangewebestörungen	1 (0,4)
Reproduktionssystem- und Bruststörungen	1 (0,4)
Summe	272 (100)

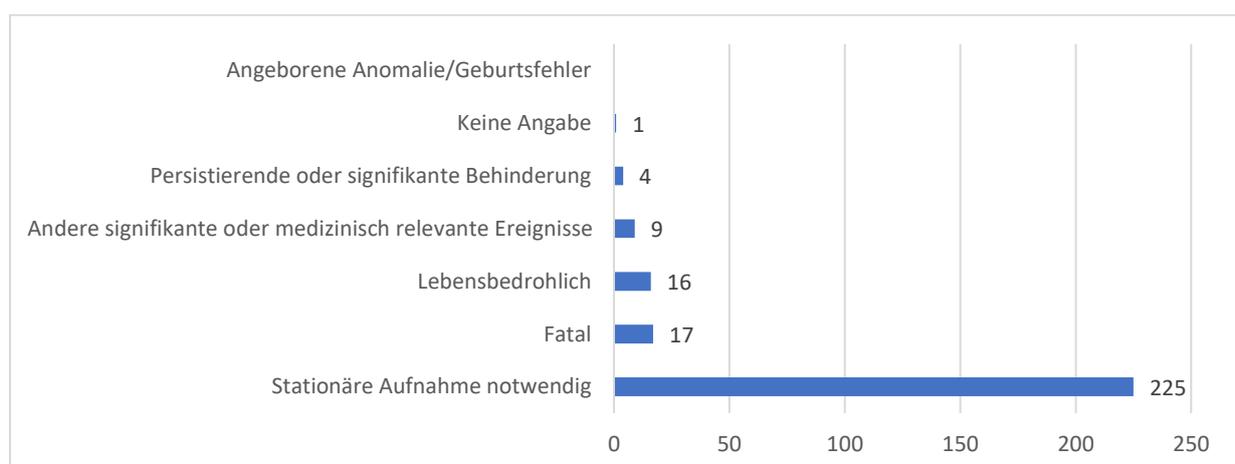
**Tabelle 11** Die Tabelle zeigt die Verteilung der schwerwiegenden unerwünschten Ereignisse nach Systemorganklassen von CTCAE V5.0.

Die Schweregradeinteilung der Ereignisse erfolgte durch lokale Studienärzte ebenfalls nach CTC-AE. 53,3 % der SUEs wiesen den Schweregrad 3 auf (n=145). Tödlichen Verlauf hatten insgesamt 8,5 % der Ereignisse (n=23). Bei 7 Ereignissen wurde keine Angabe bezüglich des Schweregrades gemacht (*siehe Tabelle 12*).

Schweregrad nach CTC-AE	n (%)
Gering	11 (4,0)
Mäßig	54 (19,9)
Schwer	145 (53,3)
Lebensbedrohlich	32 (11,8)
Tödlich	23 (8,5)
Keine Angabe	7 (2,6)

**Tabelle 12** Schweregradeinteilung der schwerwiegenden unerwünschten Ereignisse aus den beiden Studienarmen durch lokale Studienärzte nach CTC-AE.

Diese Daten beziehen sich, wie oben erläutert, auf die 272 SUE-Meldungen, die wir zur Kausalitätsprüfung und anschließend zur statistischen Auswertung ausgewählt haben. Daher zeigen die von uns erfassten Nebenwirkungen nicht das Toxizitätsprofil von Patienten in der Niveau-Studie. Darüber hinaus bewerteten wir nur schwerwiegende UEs. Deswegen war auch die Mehrheit der Ereignisse mittel- oder hochgradig nach CTCAE V5.0 (*siehe Tabelle 12*). Der weitaus häufigste Grund für die Einstufung der UEs als schwerwiegend (n=225, 83 %) war die Hospitalisierung (*siehe Abbildung 8*).



**Abbildung 8** Gründe für die Einstufung der SUEs aus den beiden Studienarmen als schwerwiegend durch lokale Studienärzte.

## 5.4 Qualität der Meldungen

Wir analysierten die Qualität der Berichte, erstattet von lokalen Studienärzten an den Sponsor. Bei 3 Meldungen (1,1 %) fehlte die Kausalitätszuordnung durch den lokalen Studienarzt. Die Information über Vorerkrankungen war fast bei der Hälfte der Meldungen nicht verfügbar (n=125, 46 %). In 40,4 % der SUEs (n=110) fehlte die Angabe über die Komedikation (abgesehen von der Tumorthherapie). Bei 13 Meldungen (4,8 %) wurden unklare medizinische Begriffe verwendet oder die relevanten Informationen in Fremdsprache vorgelegt. Das Datum der Administration des Studienmedikaments wurde lediglich bei einem SUE (0,4 %) nicht angegeben, wobei das Datum von dem Auftreten des Ereignisses war immer verfügbar. Der Beurteiler B evaluierte die Berichte zusätzlich hinsichtlich der Vollständigkeit der Fallbeschreibung. Bei mehr als einem Viertel der SUEs (n=75, 27,6 %) wurde die Beschreibung der Ereignisse als mangelhaft bewertet.

Die Schweregradbeurteilung nach CTC-AE wurde bei 7 Ereignissen (2,6 %) nicht vorgenommen. Bei 5 Ereignissen (1,8 %) wurde bezüglich des Lymphomtyps keine Angabe gemacht. Die Information über das Geschlecht und das Alter fehlte jeweils bei 2 Meldungen (0,7 %). Der Chemotherapiezyklus sowie der Grund für die Einstufung als schwerwiegend wurden nur bei einer Meldung (0,4 %) nicht angegeben. Die Information über den Studienarm, die Identifikationsnummern der Patienten und Prüfstellen war für jede Meldung verfügbar.

In den 59 Fällen (21,7 %) wurden Symptome und somit keine definierten Diagnosen als SUE berichtet. Das häufigste darunter war das Fieber (n=26, 44,1 %). Der reduzierte Allgemeinzustand war bei 12 SUEs (20,3 %) der Grund für Berichterstattung. Bei mehr als der Hälfte (n=36, 52,9 %) von den gesamten 68 SUEs aus der Gruppe Infektionen und Infestationen war die Information über Infektparameter nicht verfügbar. Aus den insgesamt 15 Ereignissen mit Pneumonie fehlte bei 6 Meldungen die Angabe über relevante bildgebende Verfahren (Computertomographie, Röntgen-Thorax und Lungensonographie).

## 5.5 Kausalitätsprüfung der schwerwiegenden unerwünschten Ereignisse

Durch lokale Studienärzte wurden insgesamt 174 Kausalitätsbewertungen der schwerwiegenden unerwünschten Ereignisse aus dem experimentellen Arm ausgeführt. Bei 3 Meldungen fehlte die Kausalitätsbewertung durch den lokalen Studienarzt. In 35,1 % der Fälle (n=61) und somit am häufigsten wurde dabei die Kategorie „Not Related“ ausgewählt. Die zweithäufigste Kategorie war mit 29,9 % (n=52) „Possible“. Am seltensten – bei lediglich 5 Ereignissen (2,9 %) - haben sich lokale Ärzte für „Certain“ entschieden. Wenn wir jeweils die positiven (sicher, wahrscheinlich und möglich) und negativen (unwahrscheinlich und nicht

möglich) Kausalitätskategorien zusammenfassen, wurde ein positiver kausaler Zusammenhang in 37,9 % der Fälle angenommen (n=66).

Der Beurteiler A hat insgesamt 94 schwerwiegende unerwünschte Ereignisse aus dem experimentellen Arm bewertet. Die beliebteste Kausalitätskategorie war mit 56,4 % „Unlikely“ (n=53). In 24,5 % der Fälle (n=23) wurde der ursächliche Zusammenhang als „möglich“ gehalten. Eine sichere Kausalität hat der Beurteiler A nur bei einem Ereignis angenommen. Eine positive Kausalität wurde in 34 % der Fälle vermutet.

Der Beurteiler B hat 117 Ereignisse aus dem experimentellen Arm zu 2 unterschiedlichen Zeitpunkten hinsichtlich der Kausalität evaluiert. Wie bei dem Beurteiler A war die Option „Unlikely“ mit 48,6 % (n=86) die häufigste Kategorie. Die zweithäufigste Wahl war mit 35,6 % (n=63) „Possible“. Die Kategorie „Certain“ wählte der Beurteiler B ebenso wie der Beurteiler A nur einmal. Ein ursächlicher Zusammenhang mit dem Studienmedikament wurde in 42,9 % der Fälle vermutet (*siehe Tabelle 13*).

Beurteiler	Ausgewählte Kausalitätskategorien					Dichotomer Outcome	
	Certain %(n)	Probable %(n)	Possible %(n)	Unlikely %(n)	Not related %(n)	Ja %	Nein %
A	1,1 (1)	8,5 (8)	24,5 (23)	56,4 (53)	9,6 (9)	34	66
B	0,6 (1)	6,8 (12)	35,6 (63)	48,6 (86)	8,5 (15)	42,9	57,1
Studienarzt	2,9 (5)	5,2 (9)	29,9 (52)	27 (47)	35,1 (61)	37,9	62,1

**Tabelle 13** Kausalitätsbewertungen von unerwünschten schwerwiegenden Ereignissen aus dem experimentellen Arm der NIVEAU Studie durch lokale Studienärzte, Beurteiler A und B.

Die Bewertungen von dem lokalen Studienarzt und dem Beurteiler A zeigten mit  $\rho=0,332$  eine geringe Korrelation. Die Korrelation zwischen den Beurteilern A und B fiel mit  $\rho=0,456$  etwa höher aus, blieb aber ebenfalls lediglich gering. Das Rangkorrelationskoeffizient für das Paar lokaler Studienarzt : Beurteiler B betrug 0,600, was einer moderaten Korrelationsstärke entspricht (*siehe Tabelle 7*). Alle Werte waren statistisch signifikant ( $P<0.001$ ).

Der Kappa-Koeffizient nach Cohen für die multivariable Likert-Skala betrug für das Paar lokaler Studienarzt : Beurteiler A nur 0,120 ( $p<0.05$ ), was einer schwachen Übereinstimmung entspricht. Für lokalen Studienarzt und Beurteiler B war die Übereinstimmung gering mit etwa höherem Kappa-Wert von 0,299 ( $P<0.001$ ). Die Beurteiler A und B zeigten ebenfalls eine geringe Übereinstimmung mit  $\kappa=0,282$  ( $P<0.001$ ). Die Tabelle 14 gibt einen Überblick über die Übereinstimmung bzw. Korrelation der Bewertungen für die Likert-Skala.

Beurteilerpaar	Spearman-Korrelation	Cohens Kappa
Studienarzt : Beurteiler A	0,332 (p<0.001)	0,120 (p<0.05)
Beurteiler A : Beurteiler B	0,456 (p<0.001)	0,282 (p<0.001)
Studienarzt : Beurteiler B	0,600 (p<0.001)	0,299 (p<0.001)

**Tabelle 14** Paarweiser Vergleich der Bewertungen anhand von dem Korrelationskoeffizienten von Spearman und Cohens Kappa für multivariable Entscheidungsform.

Aufgrund der Signifikanz für die Arzneimittelsicherheit in klinischen Prüfungen wurde die Likert-Skala auf den dichotomen Outcome reduziert (1-3 = „ja“, 4-5 = „nein“). Für die binäre Entscheidungsform wurden paarweise ungewichtete Kappa-Werte ( $\kappa$ ) ermittelt. Auch hier stimmte der Beurteiler B häufiger mit dem lokalen Studienarzt überein als der Beurteiler A und erreichte einen mittelmäßigen Übereinstimmungsgrad mit einem Kappa-Wert von 0,547 (p<0.001). Fast in 80 % der Fälle trafen sie bezüglich der Kausalität die gleiche Entscheidung. Die Übereinstimmung zwischen den beiden Beurteilern war besser als zwischen Beurteiler A und Studienärzten, blieb allerdings in beiden Fällen lediglich gering (*siehe Tabelle 15*).

Zur Bestimmung der Intrarater-Reliabilität hat der Beurteiler B 83 schwerwiegende unerwünschte Ereignisse aus dem experimentellen Arm zu dem zweiten Zeitpunkt im Abstand von mindestens 3 Monaten reevaluiert. Die ursprüngliche Kausalitätszuordnung wurde bei 13 SUEs (15,7 %) geändert. Das Korrelationskoeffizient betrug 0,912 (p<0.001), was einem sehr hohen Korrelationsgrad entspricht. Nach Dichotomisierung der Likert-Variablen berechneten wir Cohens Kappa, welches mit 0,903 (p<0,001) eine sehr gute Übereinstimmung zeigte. Die Intrarater-Reliabilität bei multivariabler Likert-Skala betrug 0,797 (p<0,001), was einer guten bis sehr guten Übereinstimmung entspricht (*siehe Tabelle 7*).

Für alle 3 Beurteilerpaare wurde zusätzlich eine paarweise exakte prozentuale Übereinstimmung (%EÜ) sowie eine extreme prozentuale Unübereinstimmung (%EU) berechnet. Eine exakte prozentuale Übereinstimmung ist definiert als die Prozentzahl der exakten Übereinstimmung in der ausgewählten Kausalitätskategorie. Fast in 60 % der Fälle wählten die Beurteiler A und B dieselbe Option in der 5-stufigen Likert-Skala (*Siehe Tabelle 16*). Für Studienärzte und den Beurteiler B war es in weniger als der Hälfte der Meldungen der Fall. Am geringsten war die %EU mit etwa 37 % für lokale Studienärzte und den Beurteiler A.

<b>a. Übereinstimmung zwischen lokalen Studienärzten und Beurteiler A</b>					
Lokaler Studienarzt					
Beurteiler A	JA	NEIN	%	Kappa	p-Wert
JA	14	16			
NEIN	15	47			
Übereinstimmung			66,3	0,227	<0,05
<b>b. Übereinstimmung zwischen lokalen Studienärzten und Beurteiler B</b>					
Lokaler Studienarzt					
Beurteiler B	JA	NEIN	%	Kappa	p-Wert
JA	51	23			
NEIN	15	85			
Übereinstimmung			78,2	0,547	<0,001
<b>c. Übereinstimmung zwischen Beurteiler A und B</b>					
Beurteiler B					
Beurteiler A	JA	NEIN	%	Kappa	p-Wert
JA	19	15			
NEIN	13	47			
Übereinstimmung			70,2	0,347	<0,001

**Tabelle 15** Die Kappa-Statistik sowie die prozentuale Übereinstimmung (%) zwischen allen 3 Beurteilerpaaren für die binäre Entscheidungsform (sicher, wahrscheinlich, möglich – Ja, unwahrscheinlich, kein Zusammenhang – Nein).

Eine extreme prozentuale Unübereinstimmung definierten wir als die Prozentzahl der Fälle, in denen Beurteiler sich für zwei nicht benachbarte Kategorien entscheiden (z. B. „Likely“ und „Unlikely“). Im Einklang mit den oben präsentierten Daten war die %UE für Studienärzte und den Beurteiler A am höchsten und betrug fast 11 %. Beurteiler A und B kamen nur bei 2 unerwünschten Ereignissen zu extremen Unübereinstimmungen (2,1 %). Die paarweise %EU lag für das Paar lokaler Studienarzt : Beurteiler B zwischen den beiden Paaren der Beurteiler und lag bei 6,3 % (siehe Tabelle 16).

Anschließend wurde für das jeweilige Paar der Beurteiler die Prozentzahl der minimalen Variationen (%MV) bestimmt, definiert als die Prozentzahl der Fälle, bei denen zwei benachbarten Kausalitätskategorien ausgewählt wurden. Aufgrund der Relevanz für die Arzneimittelsicherheit wurden daraus die Ereignisse mit dem Kategorienpaar möglich :

unwahrscheinlich ausgenommen. Die paarweise %MV war für die Beurteilerpaare Lokaler Studienarzt : Beurteiler A und lokaler Studienarzt : Beurteiler B fast identisch (29,3 % bzw. 28,7 %). Deutlich niedriger war sie mit 11,7 % für Beurteiler A und B.

Beurteilerpaar	%EÜ %(n)	%EU %(n)	%MV %(n)
Studienarzt : Beurteiler A	36,9 % (n=34)	10,9 % (n=10)	29,3 % (n=27)
Beurteiler A : Beurteiler B	58,5 % (n=55)	2,1 % (n=2)	11,7 % (n=11)
Studienarzt : Beurteiler B	48,9 % (n=65)	6,3 % (n=11)	28,7 % (n=50)

**Tabelle 16** Prozentuale Übereinstimmung zwischen Ratern. %EÜ = exakte prozentuale Übereinstimmung; %EU = exakte prozentuale Unübereinstimmung. %MV = minimale Variationen.

## 5.6 Die Evaluation der Erwartbarkeit der schwerwiegenden unerwünschten Ereignisse

Die Beurteiler A und B evaluierten 155 schwerwiegende unerwünschte Ereignisse aus beiden Studienarmen hinsichtlich der Erwartbarkeit. Der Beurteiler B bewertete die Erwartbarkeit zusätzlich bei 117 Ereignissen. Der Beurteiler B bewertete diese 117 Ereignisse zur Evaluation der Intrarater-Reliabilität zu dem zweiten Zeitpunkt im Abstand von mindestens 3 Monaten (*siehe Abbildung 6*).

Der Beurteiler A bewertete 69,7 % der Ereignisse (n=108) als erwartet und 30,3 % der Fälle (n=47) als unerwartet. Laut dem Beurteiler B war 72,1 % (n=196) der schwerwiegenden unerwünschten Ereignisse erwartet, 27,9 % der SUEs (n=76) wurden als nicht erwartet betrachtet. Die prozentuale Übereinstimmung zwischen den zwei Beurteilern betrug 87,1 %. Das Cohens Kappa belief sich auf 0,691, was auf eine gute Interrater-Reliabilität hinweist ( $p < 0,001$ ). Die Tabelle 17 gibt einen Überblick über die Bewertungen durch zwei Rater.

Beurteiler B					
Beurteiler A	JA	NEIN	%	Kappa	p-Wert
JA	99	9			
NEIN	11	36			
Übereinstimmung			87,1	0,691	p <0,001

**Tabelle 17** Die Kappa-Statistik sowie die prozentuale Übereinstimmung (%) zwischen den Beurteilern A und B bei der Evaluation der Erwartbarkeit.

Bei Reevaluation der 117 SUEs wurden bei 13 Meldungen (11,1 %) die initial getroffene Entscheidung geändert. Das Cohens Kappa zeigte mit 0,680 ( $p < 0,001$ ) eine gute Reliabilität und war somit fast genauso hoch wie beim Vergleich von zwei Experten.

Bemerkenswerterweise bewertete der Beurteiler B den kausalen Zusammenhang bei bekannten Ereignissen viermal so oft positiv (mindestens „möglich“) als bei unbekanntem SUEs (54,8 % vs. 13,7 %;  $p < 0,05$ ). Für den Beurteiler A sahen wir ein vergleichbares Verhältnis (42 % bei bekannten Ereignissen vs. 12 % bei unbekanntem SUEs;  $p < 0,001$ ). Entsprechend wurden durch Beurteiler A und B bei 86,3 % bzw. 88 % der unbekanntem Ereignisse ein ursächlicher Zusammenhang entweder als unwahrscheinlich betrachtet oder ausgeschlossen.

Beurteiler A	Positive Kausalitätskategorien % (n)	Negative Kausalitätskategorien % (n)
Erwartet	42 (29)	58 (40)
Nicht erwartet	12 (3)	88 (22)
Beurteiler B	Positive Kausalitätskategorien % (n)	Negative Kausalitätskategorien % (n)
Erwartet	54,8 (69)	45,2 (57)
Nicht erwartet	13,7 (7)	86,3 (44)

**Tabelle 18** Verteilung der ausgewählten Kausalitätskategorien nach Erwartbarkeit der Ereignisse eingestuft als solche durch den jeweiligen Beurteiler. Die Daten beziehen sich ausschließlich auf unerwünschte Ereignisse aus dem experimentellen Arm der Studie.

## 6 Diskussion

### 6.1 Kausalitätsprüfung der schwerwiegenden unerwünschten Ereignisse

Die Kausalitätsprüfung bei schwerwiegenden unerwünschten Ereignissen besitzt für Sicherheitsüberwachung in klinischen Studien eine zentrale Bedeutung. Werden die

potenziellen schädlichen Einwirkungen nicht rechtzeitig erkannt, so können die Patienten einer noch unbekanntem und sonst vermeidbaren Gefahr ausgesetzt werden.

Es wurden bereits mehrere standardisierte Kausalitätsbewertungsmethoden publiziert. Viele davon basieren auf algorithmischer Herangehensweise mit vorgegebenem Fragenkatalog und konsequenter Zuteilung der Punktwerte. Als ein Beispiel dient der von Naranjo 1981 vorgeschlagene Algorithmus(57). Der Naranjo-Algorithmus war der erste Versuch der Standardisierung von Kausalitätsbewertungen und findet bis heute in verschiedenen Einsatzgebieten als die beliebteste algorithmische Methode eine Anwendung. Aufgrund von strikt vorgeschriebenem Ablauf der Bewertung zeichnen sich die algorithmischen Methoden durch ihre gute intra- und Interrater-Reliabilität aus(57). Wegen deutlicher Prävalenz der positiven Probabilitätskategorien - insbesondere von „möglich“ - ist die Sensitivität der genannten Methode hoch. Nachteilig ist allerdings die bereits mehrmals demonstrierte niedrige Spezifität sowie geringe Flexibilität beim Einsatz in verschiedenen Gebieten(52). Eine andere ist die von dem Uppsala Monitoring Center (UMC) entwickelte Methode, welche anhand von Skalierung der Kausalitätskategorien eine effektive Grundlage für die standardisierte Evaluation der Ereignisse bietet(67). Als eine Form der globalen Introspektion ist die Reliabilität der WHO-UMC Skala allerdings niedriger(18,62). Die WHO-UMC Skala wurde zudem zur Erfassung der unbekanntem und selteneren Nebenwirkungen im Postmarketing-Bereich entwickelt und ist für den Einsatz in klinischen Studien nicht validiert(67).

Die Experteneinschätzung basiert auf individueller Evaluation der Ereignisse ohne Zuhilfenahme von standardisierten Instrumenten und wird am häufigsten - unter anderem auch in klinischen Studien - zur Kausalitätsprüfung eingesetzt(15,47). Die Umfrage von Mukherjee et al. demonstrierte, dass Experten sich häufiger auf die globale Introspektion oder sogar auf die Intuition oder „Bauchgefühl“ verlassen als auf standardisierte Methoden(56). Die Experteneinschätzung produziert aber eine niedrige Intra- und Interrater-Übereinstimmung bedingt durch subjektive Einflüsse wie Erfahrung oder Kompetenz der Beurteiler(43,46,57). Die vorhandenen Daten stammen allerdings fast ausschließlich von dem Bereich außerhalb der klinischen Prüfungen und lassen sich nicht problemlos auf klinische Studien generalisieren. Ein Grund dafür ist unter anderem, dass sie sich auf unerwünschte Arzneimittelwirkungen beziehen, wenn in klinischen Studien unerwünschte Ereignisse erfasst werden(26). Bei dem Letzteren wird der kausale Zusammenhang mit dem Studienmedikament zum Zeitpunkt der Berichterstattung nicht notwendigerweise angenommen(26). Davies et al. deuteten ebenfalls darauf hin, dass die Ereignisse aus klinischen Studien ein „Special Case“ darstellen und gesonderter Aufmerksamkeit bedürfen(18).

Diese Arbeit wurde durchgeführt, um herauszufinden, wie hoch die Übereinstimmung zwischen dem Sponsor und lokalen Studienärzten bei Bewertungen von schwerwiegenden unerwünschten Ereignissen in klinischen Studien ist. Dafür wurden die Bewertungen durch Studienärzte mit Zweitbewertungen von Sponsor (zwei unabhängige Beurteiler) verglichen, indem wir eine Interrater-Reliabilität sowie Korrelation zwischen Bewertungen ermittelten (siehe Tabelle 14 und 15). Nach unserem besten Wissen wurde die Interrater-Reliabilität zwischen Studienärzten und dem Sponsor noch nicht geprüft.

Beide Beurteiler ebenso wie lokale Studienärzte bewerteten die Mehrheit der schwerwiegenden Ereignisse als „unwahrscheinlich“ oder „kein Zusammenhang“ (62,1 % bei Studienärzten, 66 % bei dem Beurteiler A und 57,1 % bei dem Beurteiler B). Dies steht im Gegensatz zu den Daten aus dem Bereich außerhalb der klinischen Studien. Die Prädominanz der positiven Kausalitätskategorien außerhalb der klinischen Studien ist offensichtlich sowohl für einzelne Beurteiler als auch für sogenannte Goldstandards, also für konsensuelle Bewertungen durch Gruppen der Beurteiler(3). Die Häufigkeit der positiven Kausalitätsbewertungen reichte für klinische Pharmakologen bspw. von 72 % bis 84,4 % bei unerwünschten Arzneimittelwirkungen aus Krankenhausbereich und betrug 84 % beim Konsens von 5 Experten der französischen Pharmakovigilanzzentren bei der Bewertung der Ereignisse aus der Pharmakovigilanzdatenbank(3,43). Die gleiche Verteilung wird beibehalten beim Einsatz von standardisierten Methoden wie die WHO-Skala und der Naranjo-Algorithmus(1,52). Die Gründe für diese Diskrepanz dürfen in den eigentlichen Definitionen von UAW und UE liegen, denn im Letzteren wird bei der Berichterstattung die Kausalität nicht immer angenommen. Diese Unterschiede deuten darauf hin, dass die Hintergrundinformationen für UE und UAW sich maßgeblich voneinander unterscheiden. Zieht man die Tatsache in Betracht, dass verschiedene Bewertungsmethoden abhängig von Einsatzgebiet unterschiedliche Ergebnisse liefern, sollte man die bereits generierten Daten im Postmarketing-Bereich nicht pauschal auf klinische Studien generalisieren, ohne es zunächst kritisch zu hinterfragen.

Die Kausalitätszuordnungen bei unerwünschten Ereignissen aus klinischen Studien sind heterogen und werden von verschiedenen Faktoren beeinflusst. Ping-Hwa-Su und Ralph W. Stoll demonstrierten anhand der Analyse multipler Phase-I-Studien, dass die Bewertungen durch lokale Studienärzte im Großen und Ganzen den Kausalitätserwartungen entsprachen(38). Die Anzahl der positiven Kausalitätszuordnungen stieg nämlich mit zunehmender Dosierung des Studienmedikamentes an, wobei negative Bewertungen zwischen Experimentellem und Placebo-Arm gleich verteilt waren. Die Verteilung der Kausalitätszuordnungen demonstriert von Hillman et al. an dem Beispiel von 9 doppel-blinden

randomisierten Studien war mit unseren Ergebnissen vergleichbar und zeigte eine deutliche Überlegenheit negativer Kategorien in der Mehrheit der präsentierten Studien(48). Unerwünschte Ereignisse aus dem Placebo-Arm wurden zudem fast immer erwartungsgemäß häufiger als „Unlikely“ oder „Not related“ bewertet als die aus dem experimentellen Arm. Allerdings entschieden interessanterweise lokale Studienärzte in 2 Interventionsstudien bei 60 % und 62 % der UEs für mindestens „möglich“. Relativ höher Anteil von „möglich“ in diesen Studien im Vergleich zu unseren Daten lassen sich möglicherweise durch 2 Faktoren erklären. Erstens, beide Studien stellen monotherapeutische Intervention dar, im Gegensatz zu unserer Kombinationsstudie. Man kann annehmen, dass bei Kombinationstherapien Beurteiler insgesamt seltener unerwünschte Ereignisse mit Studienmedikamenten in Verbindung bringen. Die Zuordnung der Kausalität zu dem Studienmedikament ist zudem wegen häufig ähnlicher Wirkung der Arzneimittel und überlappender Toxizitäten (z. B. Zytopenien bei Chemotherapeutika) komplizierter. Aus diesem Grund schlagen George et al. sogar vor, bei Kombinationsstudien die Kausalitätsbewertung für die gesamte Therapie und nicht gesondert für das Studienmedikament vorzunehmen(30). Zweitens, von Hillman et al. wurden unerwünschte Ereignisse aller Schweregrade präsentiert. Es ist allerdings anzunehmen, dass wegen unmittelbarer Konsequenzen Beurteiler SUEs sorgfältiger bewerten und mehr Zeit in die Analyse investieren. Emanuelli demonstrierte z. B., dass die prozentuale Übereinstimmung bei den Ereignissen verschiedenen Schweregrades unterschiedlich war(22). In der Analyse der schwerwiegenden UEs durch Dorr et al. wurde bspw. analog zu unseren Daten bei der Mehrheit der schwerwiegenden Ereignisse die Kausalität für unwahrscheinlich gehalten(20). 2018 analysierte ein Auftragsforschungsinstitut 1400 SUEs aus 76 klinischen Studien und zeigte im Einklang mit unseren Daten, dass Sponsoren 67 % der Ereignisse als „Not Related“ bewerteten(44). Daher wäre eventuell weitere differenzierte Analyse notwendig, sich eine bessere Übersicht über die Kausalitätszuordnung in klinischen Studien zu verschaffen.

Der Anteil einzelner Auswahloptionen an jeweiliger positiven und negativen Kausalitätskategorien war ähnlich den bereits publizierten Daten aus klinischen Studien. Sowohl lokale Studienärzte als auch beide Beurteiler entschieden sich bei angenommenem Zusammenhang mit dem Studienmedikament weitaus häufiger für „Possible“ als für zwei andere positive Kausalitätskategorien („sicher“ und „wahrscheinlich“). Der Anteil von „Possible“ betrug nämlich 75 % bei lokalen Studienärzten, 71,9 % bei dem Beurteiler A und 82,9 % bei dem Beurteiler B. Das steht mit den Daten aus 2 Interventionsstudien analysiert durch Hillman et al. im Einklang, wo lokale Studienärzte bei angenommener Kausalität in etwa 75 % und 85 % der Fälle sich für „Possible“ entschieden(48). Für die anderen Studienarten waren die Zahlen noch höher. „Möglich“ ist die am häufigsten ausgewählte positive Kausalitätskategorie auch bei schwerwiegenden unerwünschten Ereignissen(20). Bei

Ereignissen außerhalb klinischer Studien ist die Datenlage wiederum anders. Hier nimmt der Anteil von "wahrscheinlich" und "sicher" deutlich an und übersteigt öfter den von „möglich“ unabhängig von der eingesetzten Methode(43). Eine Gruppe aus 6 Experten vermuteten bspw. bei 429 aus 500 Meldungen aus portugiesischer Pharmakovigilanzdatenbank einen kausalen Zusammenhang mit dem Arzneimittel und entschieden sich konsensuell in 57,4 Prozent der Fälle für „Certain“ oder „Likely“ und nur bei 28,5 % der Ereignisse für „Possible“(52).

Interessanterweise entschieden sich lokale Studienärzte häufiger für Extremkategorien als beide Beurteiler. Sie schließen nämlich den ursächlichen Zusammenhang mit dem Studienmedikament in 35,1 % (n=61) der Fälle komplett aus, indem sie die Kausalitätskategorie „kein Zusammenhang“ auswählten. Dabei wählten die Beurteiler A und B die Kategorie „kein Zusammenhang“ deutlich seltener und zwar lediglich in 9,6 % (n=9) bzw. 8,5 % (n=15) der Fälle aus (*siehe Tabelle 13*). Leider sind die Daten über Kausalitätszuordnungen durch den Sponsor extrem spärlich. Dorr et al. konnten allerdings zeigen, dass die Studienärzte einen signifikanten Teil für die Urteilsbildung relevanter Angaben nicht in ihren Berichten an Sponsoren übernehmen(20). Dieselben Beobachtungen machten wir auch, was nahelegt, dass den Beurteilern möglicherweise nicht genug Information zur Verfügung stand, um zur schlüssigen Entscheidung zu kommen. In den in unserer Arbeit erfassten Meldungen fehlten bspw. die Angaben über die Komedikation und Vorerkrankungen in mehr als 40 % der Meldungen. Die Beschreibungen der Ereignisse befand der Beurteiler B in fast 30 % der Fälle für nicht ausreichend.

Die Übereinstimmung zwischen den Bewertungen war mit den bereits gezeigten Daten aus unterschiedlichen Bereichen der Pharmakovigilanz vergleichbar. Zur Übereinstimmung gelangen am häufigsten lokale Studienärzte und der Beurteiler B. Sie zeigten nämlich eine moderate Korrelation ( $\rho=0,600$ ;  $p<0.001$ ) und eine mittelmäßige Übereinstimmung bei binärer Entscheidungsform ( $\kappa=0,547$ ;  $p<0.001$ ). Bei Bewertungen nach Likert-Skala erreichte die Interrater-Reliabilität höchstens eine geringe Übereinstimmung. Für das Paar lokaler Studienarzt : Beurteiler A war sie sogar schwach ( $\kappa=0,120$ ;  $p<0.05$ ). Der Beurteiler B wies hier mit Studienärzten sowie mit dem Beurteiler A den ähnlichen Grad der Übereinstimmung auf mit einem  $\kappa$  von 0,299 bzw. 0,282 ( $p<0.001$ ).

Naranjo et al. demonstrierten in ihrer Originalstudie ebenfalls nur eine geringe Übereinstimmung für globale Introspektion zwischen Experten bei der Bewertung von UAW mit einem Kappa-Wert von 0,21 bis 0,37(57). Auch die prozentuale Übereinstimmung war mit von uns gezeigten Ergebnissen fast identisch und reichte von 41 % bis 57 %, während sie in unserem Fall von 36,9 % bis 58,5 % schwankte. Arimone et al. untersuchten die

Übereinstimmung zwischen 5 Experten aus französischen Pharmakovigilanzzentren mit mindestens 10-jähriger Erfahrung in der Bewertung von Spontanmeldungen und zeigten ebenso eine geringe globale Übereinstimmung mit einem Kappa-Wert von 0,20(4).

Die Daten über die Übereinstimmung zwischen lokalen Studienärzten und Sponsoren bei der Bewertung unerwünschter Ereignisse aus klinischen Studien sind sehr spärlich. In der Analyse von mehr als 1400 SUEs aus 76 Studien durch ein Auftragsforschungsinstitut wurde demonstriert, dass Sponsoren häufiger schwerwiegende Ereignisse mit dem Studienmedikament in Verbindung brachten als Studienärzte aus Krankenhäusern(44). Dieses Verhältnis konnten wir für beide Beurteiler des Sponsors reproduzieren. Sie hielten nämlich in 66 % bzw. 57,1 % der Fälle den Kausalzusammenhang für unwahrscheinlich, für lokale Studienärzte waren es 62,1 %. Kosov et al. evaluierten Auswertungen von 50 schwerwiegenden unerwünschten Ereignissen aus 3 klinischen Studien Phase II und III(45). Die prozentuale Übereinstimmung zwischen Studienärzten und den jeweiligen Sponsoren erreichte für das dichotome Outcome 96 % (48/50). In unserem Fall betrug sie für die Beurteiler A und B jeweils 66,3 % und 78,2 %. Was für die globale Introspektion derartig ungewöhnlich hohe Übereinstimmung bewirkt hat, ist nicht klar, Autoren nehmen keinen Bezug auf die präsentierten Daten und führen keine Reliabilitätsprüfung durch. Zusätzlich wurde die Übereinstimmung zwischen Studienärzten und 4 unabhängigen Experten geprüft. Für 3 von 4 Experten war sie im Bereich von 70 %-74 % vergleichbar mit unseren Ergebnissen (66 % und 78 % jeweils für Beurteiler A und B), der 4. Experte mit geringerer Erfahrung zeigte eine deutlich niedrigere Übereinstimmung mit 44 %. Bemerkenswerterweise wies in unserem Fall der Beurteiler mit geringerer Erfahrung (der Beurteiler B) höhere Übereinstimmung mit Studienärzten auf. Der Vergleich von Beurteilern anhand einer angemessenen statistischen Methode wurde allerdings auch hier leider nicht vorgenommen.

Wir demonstrierten ferner eine sehr gute Intrarater-Reliabilität für den Beurteiler B (Kappa 0,903,  $p < 0,001$ ) vergleichbar mit der Intrarater-Reliabilität beim Einsatz algorithmischer Methode, berichtet von Naranjo et al.(57). Auch die Korrelation zwischen zwei Bewertungen war mit einem  $\rho$  von 0,912 sehr stark. Für die Retest-Reliabilität gibt es insbesondere aus dem Bereich der klinischen Studien noch keine Daten zum Vergleich. Für die einstimmig als unzuverlässig anerkannte Methode lässt sich derartig hohe Konsistenz der Bewertungen schwierig erklären und deutet mglw. darauf hin, dass die Urteilsbildung in Bezug auf die Kausalität bei globaler Introspektion zwar subjektiv sein kann, die Entscheidungen basieren aber auf objektiven und konsistenten Faktoren. Die Ergebnisse sollten allerdings nicht überschätzt werden. Die Intrarater-Reliabilität bei globaler Introspektion bedarf unserer Ansicht nach mehr Aufmerksamkeit.

Eines sollten wir noch abschließend betonen: Die Daten über die Interrater-Reliabilität von Experteneinschätzung stammen nicht nur aus sehr unterschiedlichen Bereichen der Pharmakovigilanz, sie sind auch aus dem statistischen Gesichtspunkt oft vortäuschend. Die ungewichtete Kappa-Statistik bei der multivariablen Likert-Skala drückt die tatsächliche Übereinstimmung zwischen Experten nur ungenügend aus, denn die Unterschiede zwischen den Kategorien „sicher“ und „wahrscheinlich“ sind insbesondere im regulatorischen Sinne nicht mit dem Unterschied zwischen den Kategorien sicher und kein Zusammenhang gleichzusetzen. Die Daten bei binärer Entscheidungsform sind aber auch aus dem Bereich außerhalb der klinischen Studien spärlich. Weitere Analysen der Übereinstimmung zwischen Experten in diesem Setting sind essenziell, in erster Linie die Ursachen von unterschiedlichen Bewertungen feststellen und diese minimieren zu können.

## **6.2 Möglichkeiten zur Besserung der Übereinstimmung zwischen lokalen Studienärzten und Sponsor**

Die grundlegende Schwierigkeit bei der Bewertung unerwünschter Ereignisse besteht darin, dass ein sicherer Nachweis der Kausalität nicht möglich ist. Dementsprechend ist die Prüfung der Validität der Bewertungsmethoden ein Problem, was nicht zu lösen zu sein scheint. Die Prüfung der Kausalität ähnelt daher in der Praxis im Wesentlichen einer Differenzialdiagnostik am Patientenbett. Aus diesem Grund ist es auch verständlich, dass Experten oft zu unterschiedlichen Schlussfolgerungen kommen. Somit stellt sich auch die Frage, welcher Übereinstimmungsgrad sollte dann - wenn überhaupt - angestrebt werden? Hier sollten wir noch eins anmerken: Selbst eine perfekte Übereinstimmung bedeutet nicht, dass die ausgewählte Option richtig ist. Die Möglichkeiten zur Besserung der Interrater-Reliabilität müssen nur in diesem Kontext angesehen werden.

Wie wir demonstrieren konnten, ist die in CRFs dokumentierte Information in einem beträchtlichen Anteil der Meldungen mangelhaft (*siehe Abschnitt 5.6*). Dorr et al. zeigten zusätzlich, dass die Angaben, die Kausalitätszuordnungen durch Studienärzte untermauerten, viel häufiger in der primären Datenquelle aus Studienzentren zu finden waren als in erstatteten Meldungen an den Sponsoren(20). Selbstverständlich kann man davon ausgehen, dass eine niedrige Berichtsqualität die Übereinstimmung zwischen dem Berichterstatter und dem Zweitbewerter negativ beeinflusst. Das könnte auch erklären, warum lokale Studienärzte in unserem Fall häufiger als die Beurteiler von dem Sponsor sich für Extremkategorien entschieden. Kausalitätszuordnungen aus dem gleichen Informationsstand müssten häufiger übereinstimmen.

Ein anerkannter Standard für die Abstufung der Kausalität existiert nicht. Häufig wird eine Antwortskala mit mehreren Auswahloptionen bevorzugt, darunter auch von Bundesoberbehörden BfArM und PEI(9,58)(siehe Tabelle 1). „Möglich“ ist standardmäßig bei der Sicherheitsanalyse in klinischen Studien für die Zuordnung der Ereignisse zu dem Prüfmedikament im regulatorischen Sinne ausreichend. Die durch uns sowie in früheren Analysen demonstrierte Häufigkeit dieser Kausalitätskategorie bedarf unserer Ansicht nach der besonderen Aufmerksamkeit. Begriffliche Unterschiede zwischen einzelnen Kausalitätskategorien sind nicht eindeutig und gerade für „möglich“ ist das Risiko unterschiedlicher Interpretation zwischen Beurteilern am höchsten, denn alle Kategorien außer „Not Related“ fallen immer noch unter „möglich“ (das Medikament ist „unwahrscheinlich“ das Ereignis ausgelöst zu haben, die Kausalität ist aber immer noch „möglich“, weil es nicht ausgeschlossen – also unmöglich – ist). Bei Experten mit unterschiedlichen sprachlichen Hintergründen sind diese feinen semantischen Differenzen noch ausgeprägter. Darauf wird auch von der CIOMS Arbeitsgruppe VI hingedeutet(15). Um diesen Störfaktor zu vermeiden, erscheint eine binäre Kausalitätsbewertung angemessener. George et al. empfehlen eine wahrscheinlichkeitsbasierte Wortwahl wie z. B. “more likely related to study drug (than other causes)” und “more likely related to other causes (than study drug)”(30).

Bei Experteneinschätzung ist eine umfassende Analyse aller zur Verfügung stehenden Informationen durch Beurteiler unerlässlich. Diese Informationen bilden eine gewisse intrinsische „Evidenz“ für das gefällte Urteil. Zur Bewertung der Spontanmeldungen wurde von dem WHO-Monitoring-Zentrum in Uppsala eine Wahrscheinlichkeitsskala vorgeschlagen. Die WHO-UMC Skala gewährleistet einerseits, dass alle Kriterien der intrinsischen Evidenz (siehe Abschnitt 3.2.1) berücksichtigt werden. Andererseits ermöglicht sie ein standardisiertes Vorgehen bei globaler Introspektion, was verständlicherweise zuverlässiger erscheint als das Gegenüberstellen der Argumente „im Kopf“. Die WHO-UMC Skala ist allerdings eine Likert-Skala und wurde nicht für den Einsatz in klinischen Studien konzipiert. Eine adaptierte Wahrscheinlichkeitsskala, konzipiert auf der Grundlage der WHO-UMC Skala, wird in der Tabelle 19 präsentiert. Die Gewichtung der Fragen ist entsprechend abgeschafft. Der einfache Aufbau der Skala ermöglicht ggf. die Integration im Meldebogen bzw. im Bewertungsbogen.

<p><b>Der Zusammenhang mit dem Studienmedikament ist wahrscheinlicher (als mit anderen Medikamenten oder Krankheit) wenn:</b></p>
<ul style="list-style-type: none"> <li>▪ Zeitlicher Zusammenhang mit dem Ereignis ist gegeben</li> <li>▪ Studienmedikament bietet mindestens genauso plausible Erklärung wie die Krankheit oder andere Medikamente</li> <li>▪ Das Ereignis ist pharmakologisch und phänomenologisch definiert</li> <li>▪ Rechallenge zufriedenstellend, wenn durchgeführt</li> <li>▪ Dechallenge ist zufriedenstellend, wenn durchgeführt</li> </ul>
<p><b>Der Zusammenhang mit anderen Medikamenten oder Krankheit ist wahrscheinlicher (als mit dem Studienmedikament) wenn:</b></p>
<ul style="list-style-type: none"> <li>▪ Zeitlicher Zusammenhang mit dem Auftreten des Ereignisses und der Medikamenteneinnahme macht den Kausalzusammenhang unwahrscheinlich</li> <li>▪ Krankheit oder andere Medikamente liefern bessere Erklärung als das Studienmedikament</li> </ul>

**Tabelle 19** Kausalitätsbewertungsskala mit wahrscheinlichkeitsbasierten Kausalitätskategorien für die binäre Entscheidungsform abgeleitet von der WHO-UMC Skala.

### 6.3 Die Evaluation der Erwartbarkeit der schwerwiegenden unerwünschten Ereignisse

Eins der kritischen Momente beim Sicherheitsmonitoring in klinischen Studien ist die Entscheidung darüber, ob das aufgetretene schwerwiegende unerwünschte Ereignis erwartet war. Diese Entscheidung basiert darauf, ob das Ereignis in der Art und Ausprägung bereits in der Referenz-Sicherheitsinformation (RSI) erfasst wurde(26). Als RSI kann für die bereits zugelassenen Medikamente die Fachinformation dienen, für Medikamente, die noch keine Zulassung auf Landesebene haben, wird als Referenz dagegen die Investigator's Brochure genommen(10,15).

Wird bei einem schwerwiegenden Ereignis - welches noch nicht in der entsprechenden RSI beschrieben ist - der Kausalzusammenhang mit dem Studienmedikament angenommen, so handelt es sich mglw. um eine noch unbekannte Nebenwirkung, die die Sicherheit der Studienteilnehmer gefährden könnte. Über diese Ereignisse müssen unverzüglich, spätestens jedoch nach 15 Tagen die zuständige Ethik-Kommission, die zuständige Bundesbehörde (BfArM/PEI) sowie die Studienärzte beteiligter Zentren informiert werden(26). Eine korrekte Evaluation der Erwartbarkeit verhindert zudem die Überflutung mit SUSARs ohne Sicherheitsrelevanz für Prüfungsteilnehmer.

Die Evaluation von Erwartbarkeit umfasst im Wesentlichen die Prüfung von s. g. „Listedness“, was grundsätzlich bedeutet, dass es überprüft werden muss, ob das Ereignis bereits in der RSI aufgelistet ist(26). Wir prüften die Übereinstimmung zwischen zwei Beurteilern, indem wir

die Beurteilungen von Erwartbarkeit bei schwerwiegenden unerwünschten Ereignissen aus der NIVEAEU Studie verglichen. Der Vorgang der Evaluation von Erwartbarkeit lässt auf den ersten Blick kaum Raum für die unterschiedliche Interpretation. Dennoch kamen 2 Beurteiler in fast 13 % der Fälle zu unterschiedlichen Ergebnissen. Noch auffallend ist unsere Analyse der Intrarater-Reliabilität. Der Beurteiler B hat die 117 SUEs zum zweiten Zeitpunkt reevaluiert und in mehr als 11 % der Fälle die ursprüngliche Entscheidung geändert. Das entspricht einem Kappa-Wert von 0,680 ( $p < 0,001$ ). Die Interrater-Reliabilität war somit sogar etwas höher, Cohens Kappa belief sich nämlich auf 0,691 ( $p < 0,001$ ).

Die EU-Verordnung definiert, welche Ereignisse als unerwartet bezeichnet werden müssen(26). Die Definition der Erwartbarkeit ist allerdings nicht unkompliziert. Es gibt dabei erstaunlich wenig Daten darüber, wie effektiv die übliche Praxis der Evaluation von Erwartbarkeit ist. Der Kappa-Wert für die Übereinstimmung der Bewertungen von Erwartbarkeit wurde noch nie publiziert. Jarow et al. führten 2016 in der Datenbank der FDA ein formales Audit von SUSAR-Meldungen durch, welche durch private Sponsoren gemeldet wurden. Sie stellten fest, dass 54 % der Ereignisse in der Referenz-Sicherheitsinformation bereits erwähnt und somit nicht unbekannt waren(41). Den Autoren bleibt es allerdings unklar, welche Art der RSI von den jeweiligen Sponsoren genutzt wurde und geben selbst ebenfalls nicht an, welche Form der RSI im Audit als Referenz diente. Kosov et al. zeigten, dass bei der Evaluation von 50 schwerwiegenden unerwünschten Ereignissen Beurteiler in 12 % der Fälle zu unterschiedlichen Schlussfolgerungen kamen(45). Die Kappa-Statistik wurde hier nicht ermittelt. Die prozentuale Übereinstimmung ist allerdings fast identisch mit unseren Daten (48 % vs. 47,1 %).

Wir setzten uns mit den SUEs auseinander, die bei der Bewertung von Erwartbarkeit für Unstimmigkeit zwischen 2 Beurteilern sorgten. Es fielen 4 mögliche Störfaktoren auf:

1. Die EU-Verordnung sowie die ICH-Leitlinie besagen, dass Ereignisse, die spezifischer sind als die in der RSI, als unerwartet bezeichnet werden müssen(25,26). Ein umgekehrter Fall ist aber ebenfalls möglich. Bei dem SUE Nummer 19 wurde z. B. eine gastrointestinale Blutung berichtet. In der RSI ist die Ulkuskrankheit aufgelistet – die häufigste Blutungsursache. Ist das Ereignis erwartet? Die 2 Beurteiler sahen es unterschiedlich. Ein weiteres Beispiel aus unserer Studie stellt Sepsis dar. Sepsis ist ein Sammelbegriff, der nicht in der RSI von Nivolumab aufgelistet ist. Infektionen der unterschiedlichen Lokalisationen, die mit Sepsis einhergehen oder in Sepsis übergehen können, sind aber in der RSI erwähnt. 3 Fälle mit Sepsis wurden von Beurteilern unterschiedlich bewertet.

2. Obwohl die Feststellung einer spezifischen Diagnose durch ICH Leitlinie ermutigt wird, sind Studienärzte grundsätzlich erlaubt, Symptome oder Laborbefunde als schwerwiegende Ereignisse zu melden(25). Dies ist insbesondere der Fall, wenn noch keine Diagnose gestellt ist und die Meldefrist eingehalten werden muss. So umfangreich können die RSIs natürlich nicht sein. Vier Ereignisse, die als Symptome oder Krankheitszeichen berichtet wurden, wurden durch Beurteiler unterschiedlich interpretiert. Als Beispiel kann das SUE 100 dienen. Berichtet wurde hier über die akute respiratorische Insuffizienz, die nicht in der Fachinformation zu finden war. Es sind aber andere Erkrankungen aufgelistet, die respiratorische Insuffizienz auslösen können, darunter auch die Pneumonie. Das Gleiche betrifft auch den „reduzierten Allgemeinzustand“ - das zweithäufigste Ereignis aus der Gruppe „allgemeine Erkrankungen und Beschwerden am Verabreichungsort“ (*siehe Abschnitt 5.5*).
3. Die ICH-GCP-Leitlinie und die EU-Verordnung definieren nicht, ob verschiedene Ereignisse derselben Art als unerwartet betrachtet werden sollen. In der RSI von Nivolumab sind z. B. einige Infektionen aufgelistet. Sind dann alle Infektionen in regulatorischem Sinne erwartet? Aus dem medizinischen Gesichtspunkt erscheint es zumindest bei Medikamentenstudien sinnvoll. Bei mindestens 4 Ereignissen könnte dieser Faktor der Grund für unterschiedliche Bewertung gewesen sein.
4. Manchmal werden unbekannte Ereignisse von bekannten Ereignissen hervorgerufen. Das war vermutlich bei dem SUE 148 der Fall, wo der Patient mit schwerer Thrombozytopenie unter isolierter Hämaturie litt. Thrombozytopenie ist für Nivolumab ein erwartetes Ereignis, Hämaturie – nicht. Der Beurteiler A bewertete das Event als erwartet, der Beurteiler B – als unerwartet.

Diese Beispiele zeigen, dass die Beurteilung der Erwartbarkeit komplizierter sein kann als nur das „Abgleichen“ der Liste von bekannten Ereignissen in der RSI. Es erscheint zumindest plausibel, dass die Klärung dieser Unsicherheiten im Vorfeld durch Sponsor (z. B. in der Investigator's Brochure) zur einheitlichen und konsistenten Evaluation der Erwartbarkeit beitragen kann. Für die meisten der SUEs sahen wir allerdings keinen sicheren Grund für unterschiedliche Bewertungen. Sie wurden entweder als erwartet bewertet, obwohl sie nicht aufgelistet waren oder – als unerwartet, obwohl sie in der Fachinformation erwähnt waren. Bei Intrarater-Reliabilitätsprüfung wurden wiederum dieselben Ereignisse zu dem zweiten Zeitpunkt unterschiedlich bewertet, die beim Vergleich von 2 Beurteilern für Unstimmigkeit sorgten. Bei 2 Meldungen mit Diagnose Sepsis wurde z. B. die ursprüngliche Entscheidung geändert.

Bei der Evaluation der Erwartbarkeit wurde durch beide Beurteiler vom Sponsor der NIVEAU-Studie die Fachinformation als Referenz genommen. Für Arzneimittel, die auf Landesebene nicht zugelassen sind, muss die Investigator's Brochure genutzt werden(25). Investigator's Brochure ist eine umfassende Zusammenstellung aller klinischen und nicht klinischen Daten über das Prüfpräparat. Die Fachinformation ist dagegen eine kompaktere Darstellung der sicherheitsrelevanten Daten zu dem Medikament. Vor diesem Hintergrund erscheint es interessant, ob die Art des Referenz-Dokuments die Übereinstimmung bei der Bewertung von Erwartbarkeit beeinflussen könnte. Kosov et al. berichten über prozentuale Übereinstimmung von 48 % für die Investigator's Brochure(45). Bei der Fachinformation in der NIVEAU-Studie war sie mit 47,1 % vergleichbar. Die Fallzahl bei Kosov et al. war allerdings deutlich geringer (n=50). Zum besseren Verständnis der möglichen Gründe für Unstimmigkeiten bei der Evaluation der Erwartbarkeit sind mehr Daten erforderlich.

## 7 Literaturverzeichnis

1. Acharya TA, Trivedi MD, Joshi KJ, Chhaiya SB, Mehta DS. A study of agreement between WHO-UMC causality assessment system and the Naranjo algorithm for causality assessment of adverse drug reactions observed in medical ICU of a tertiary care teaching hospital. *Biomed Pharmacol J.* 2020;13(1):79–83.
2. Anderson N, Borlak J. Correlation versus causation? pharmacovigilance of the analgesic flupirtine exemplifies the need for refined spontaneous adr reporting. *PLoS One.* 2011;6(10).
3. Arimone Y, Bégaud B, Miremont-Salamé G, Fourier-Réglat A, Molimard M, Moore N, et al. A new method for assessing drug causation provided agreement with experts' judgment. *J Clin Epidemiol.* 2006;59(3):308–14.
4. Arimone Y, Bégaud B, Miremont-Salamé G, Fourier-Réglat A, Moore N, Molimard M, et al. Agreement of expert judgment in causality assessment of adverse drug reactions. *Eur J Clin Pharmacol.* 2005;61(3):169–73.
5. Arimone Y, Miremont-Salamé G, Haramburu F, Molimard M, Moore N, Fourier-Réglat A, et al. Inter-expert agreement of seven criteria in causality assessment of adverse drug reactions. *Br J Clin Pharmacol.* 2007;64(4):482–8.
6. Behera SK, Das S, Xavier AS, Velupula S, Sandhiya S. Comparison of different methods for causality assessment of adverse drug reactions. *Int J Clin Pharm.* 2018;40(4):903–10.
7. Bishop MR, Dickinson M, Purtill D, Barba P, Santoro A, Hamad N, et al. Second-line tisagenlecleucel or standard care in aggressive B-cell lymphoma. *N Engl J Med.* 2022;386(7):629–39.
8. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). Pharmakovigilanz bei Älteren – Herausforderungen der Kausalitätsbewertung anhand von Fallbeispielen aus der Nebenwirkungsdatenbank des BfArM. *Bull Zur Arzneimittelsicherheit* [Internet]. 2015 [cited 2024 Jan 11];(2):21–34. Available from: [http://www.bfarm.de/SharedDocs/Downloads/DE/Arzneimittel/Pharmakovigilanz/Bulletin/2015/4-2015.pdf?\\_\\_blob=publicationFile&v=5](http://www.bfarm.de/SharedDocs/Downloads/DE/Arzneimittel/Pharmakovigilanz/Bulletin/2015/4-2015.pdf?__blob=publicationFile&v=5)
9. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). Bericht Über Unerwünschte Arzneimittelwirkungen (Auch Verdachtsfälle) [Internet]. [cited 2024 Jan 11]. Available from: <https://www.bfarm.de/SharedDocs/Formulare/DE/Arzneimittel/Pharmakovigilanz/aa-uaw-melde-bogen.html>
10. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), Paul-Ehrlich-Institut (PEI). Pharmakovigilanz in klinischen Prüfungen. *Bull Zur Arzneimittelsicherheit*

- [Internet]. 2016 [cited 2024 Jan 11];(2):26–30. Available from:  
[http://www.bfarm.de/SharedDocs/Downloads/DE/Arzneimittel/Pharmakovigilanz/Bulletin/2015/4-2015.pdf?\\_\\_blob=publicationFile&v=5](http://www.bfarm.de/SharedDocs/Downloads/DE/Arzneimittel/Pharmakovigilanz/Bulletin/2015/4-2015.pdf?__blob=publicationFile&v=5)
11. Bundesinstitut für Arzneimittel und Medizinprodukte, Paul-Ehrlich-Institut. 6. Bekanntmachung zur Anzeige von Nebenwirkungen und Arzneimittelmisbrauch nach § 63b Absatz 1 bis 9 des Arzneimittelgesetzes (AMG). 2010;
  12. Bundesministerium für Gesundheit und Soziale Sicherung. Verordnung über die Anwendung der Guten Klinischen Praxis bei der Durchführung von klinischen Prüfungen mit Arzneimitteln zur Anwendung am Menschen (GCP-Verordnung - GCP-V) § 3 Begriffsbestimmungen [Internet]. Vol. 42, Bundesgesetzblatt. Bonn; 2004 [cited 2024 Jan 11]. p. 2081–91. Available from:  
[https://www.bgbl.de/xaver/bgbl/start.xav?start=/\\*%5B@attr\\_id=%27bgbl104s3235.pdf%27%5D#\\_\\_bgbl\\_\\_%2F%2F%5B%40attr\\_id%3D%27bgbl104s2081.pdf%27%5D\\_\\_1655231725808](https://www.bgbl.de/xaver/bgbl/start.xav?start=/*%5B@attr_id=%27bgbl104s3235.pdf%27%5D#__bgbl__%2F%2F%5B%40attr_id%3D%27bgbl104s2081.pdf%27%5D__1655231725808)
  13. Chan YH. Biostatistics 104: Correlational Analysis. Singapore Med J. 2005;44(12):614–9.
  14. Chau I, Watkins D, Cunningham D. Gemcitabine and its combinations in the treatment of malignant lymphoma. Clin Lymphoma. 2002;3(2):97–104.
  15. CIOMS. Management of Safety Information from Clinical Trials [Internet]. Geneva; 2005 [cited 2024 Jan 11]. Available from: [https://cioms.ch/wp-content/uploads/2017/01/Mgment\\_Safety\\_Info.pdf](https://cioms.ch/wp-content/uploads/2017/01/Mgment_Safety_Info.pdf)
  16. Cleeland CS, Zhao F, Chang VT, Sloan JA, O'Mara AM, Gilman PB, et al. The symptom burden of cancer: Evidence for a core set of cancer-related and treatment-related symptoms from the Eastern Cooperative Oncology Group Symptom Outcomes and Practice Patterns study. Cancer. 2013;119(24):4333–40.
  17. Coiffier B, Thieblemont C, Van Den Neste E, Lepage G, Plantier I, Castaigne S, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: A study by the Groupe d'Etudes des Lymphomes de l'Adulte. Blood. 2010;116(12):2040–5.
  18. Davies EC, Rowe PH, James S, Nickless G, Ganguli A, Danjuma M, et al. An investigation of disagreement in causality assessment of adverse drug reactions. Pharmaceut Med. 2011;25(1):17–24.
  19. Doherty MJ. Algorithms for assessing the probability of an Adverse Drug Reaction. Vol. 2, Respiratory Medicine CME. 2009. p. 63–7.
  20. Dorr DA, Burdon R, West DP, Lagman J, Georgopoulos C, Belknap SM, et al. Quality of reporting of serious adverse drug events to an institutional review board: A case

- study with the novel cancer agent, imatinib mesylate. *Clin Cancer Res.* 2009;15(11):3850–5.
21. Eaton A, Iasonos A, Gounder MM, Pamer EG, Drilon A, Vulih D, et al. Toxicity Attribution in Phase I Trials: Evaluating the Effect of Dose on the Frequency of Related and Unrelated Toxicities. *Clin Cancer Res.* 2016;22(3):553–9.
  22. Emanuel A. A Simple Algorithm for Assessing Causality of Adverse Reactions Special Workshop—Industrial. *Drug Inf J.* 1984;18(3–4):303–6.
  23. Enrico D, Waisberg F, Burton J, Mandó P, Chacón M. Analysis of adverse events attribution and reporting in cancer clinical trials: A systematic review. *Crit Rev Oncol Hematol.* 2021;160:103296.
  24. European Medicines Agency (EMA). Guideline For Good Clinical Practice E6(R2). *Comm Hum Med Prod [Internet].* 2018 [cited 2024 Jan 11];6(December 2016):1–68. Available from: [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf)
  25. European Medicines Agency (EMA). ICH Harmonised Tripartite Guideline E2A: Definitions and Standards for Expedited Reporting. 1995 [cited 2024 Jan 11]; Available from: [https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use\\_en-15.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-15.pdf)
  26. European Parliament and the Council of the European Union. Official Journal of the European Union [Internet]. REGULATION (EU) No 536/2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. 2014 [cited 2024 Jan 11]. p. 1–76. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0536&from=EN>
  27. Faivre S, Raymond E, Woynarowski JM, Cvitkovic E. Supraadditive effect of 2',2'-difluorodeoxycytidine (gemcitabine) in combination with oxaliplatin in human cancer cell lines. *Cancer Chemother Pharmacol.* 1999;44(2):117–23.
  28. Fiedler C, Raddatz B. Study Nurse / Studienassistentz. Sicherheit in klinischen Prüfungen. Springer Verlag; 2015. 71–80 p.
  29. Gallagher RM, Kirkham JJ, Mason JR, Bird KA, Williamson PR, Nunn AJ, et al. Development and inter-rater reliability of the Liverpool adverse drug reaction causality assessment tool. *PLoS One.* 2011;6(12).
  30. George GC, Barata PC, Campbell A, Chen A, Cortes JE, Hyman DM, et al. Improving attribution of adverse events in oncology clinical trials. *Cancer Treat Rev.* 2019;76(June):33–40.
  31. German Lymphoma Alliance. Protokollsynopse DSHNHL 2015-1 zu Protokoll Version 06.0-F vom 21.12.2020 [Internet]. 2020 [cited 2022 Aug 29]. Available from:

- [https://www.german-lymphoma-alliance.de/media/public/C2DC5D65-996F-66D6-08E1-6AC9B0F35197/deutsche-Uebersetzung-Protokollsynopse-DSHNHL-2015-1\\_V06.0\\_F\\_clean.pdf?ts=1624267984](https://www.german-lymphoma-alliance.de/media/public/C2DC5D65-996F-66D6-08E1-6AC9B0F35197/deutsche-Uebersetzung-Protokollsynopse-DSHNHL-2015-1_V06.0_F_clean.pdf?ts=1624267984)
32. Gisselbrecht C, Glass B, Mounier N, Gill DS, Linch DC, Trneny M, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol*. 2010;28(27):4184–90.
  33. El Gnaoui T, Dupuis J, Belhadj K, Jais JP, Rahmouni A, Copie-Bergman C, et al. Rituximab, gemcitabine and oxaliplatin: An effective salvage regimen for patients with relapsed or refractory B-cell lymphoma not candidates for high-dose therapy. *Ann Oncol*. 2007;18(8):1363–8.
  34. Held G. Improvement of Outcome in Elderly Patients or Patients not eligible for high-dose chemotherapy with Aggressive Non-Hodgkin Lymphoma in first Relapse or Progression by adding Nivolumab to Gemcitabine, Oxaliplatin plus Rituximab in case of B-cell lymphoma. *Studienprotokoll*. 2018;
  35. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med*. 1965;58(5):295–300.
  36. Hillman SL, Mandrekar SJ, Bot B, DeMatteo RP, Perez EA, Ballman K V., et al. Evaluation of the value of attribution in the interpretation of adverse event data: A North Central Cancer Treatment Group and American College of Surgeons Oncology Group investigation. *J Clin Oncol*. 2010;28(18):3002–7.
  37. Hsu PH, Stoll RW. Causality assessment of adverse events in clinical trials: II. An algorithm for drug causality assessment. *Drug Inf J*. 1993;27(2):387–94.
  38. Hsu PH, Stoll RW. Causality assessment of adverse events in clinical trials: I. How good is the investigator drug causality assessment? *Drug Inf J*. 1993;27(2):377–85.
  39. Hutchinson TA. Computerized bayesian ADE assessment. *Drug Inf J*. 1991;25(2):235–41.
  40. Iasonos A, Gounder M, Spriggs DR, Gerecitano JR, Yman DM, Zohar S, et al. The impact of non-drug-related toxicities on the estimation of the maximum tolerated dose in phase I trials. *Clin Cancer Res*. 2013;18(19):5179–87.
  41. Jarow JP, Casak S, Chuk M, Ehrlich LA, Khozin S. The majority of expedited investigational new drug safety reports are uninformative. *Clin Cancer Res*. 2016;22(9):2111–3.
  42. Kamdar M, Solomon SR, Arnason J, Johnston PB, Glass B, Bachanova V, et al. Lisocabtagene maraleucel versus standard of care with salvage chemotherapy followed by autologous stem cell transplantation as second-line treatment in patients with relapsed or refractory large B-cell lymphoma (TRANSFORM): results from an interim analysis. *Lancet*. 2022;399(10343):2294–308.

43. Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. Vol. 11, *European Journal of Clinical Pharmacology*. 1977. p. 75–8.
44. Kopetskie H. Site Investigator vs. Sponsor SAE Causality: Are they different? [Internet]. Contact Research Organisation Rho. 2018 [cited 2022 Sep 2]. Available from: <https://www.rhoworld.com/site-investigator-vs-sponsor-sae-causality-are-they-different/#:~:text=When a serious adverse event,the drug caused the event>.
45. Kosov M, Maximovich A, Riefler J, Batson E, Dignani MC, Belotserkovskiy M. Interexpert agreement on adverse events' evaluation. *Appl Clin Trials*. 2016;25(8):43–8.
46. Kramer MS. Difficulties in assessing the adverse effects of drugs. *Br J Clin Pharmacol*. 1981;11(S1):105–10.
47. Kramer MS. Assessing Causality of Adverse Drug Reactions: Global Introspection and Its Limitations. *Drug Inf J*. 1986;20(4):433–7.
48. Le-Rademacher J, Hillman SL, Meyers J, Loprinzi CL, Limburg PJ, Mandrekar SJ. Statistical controversies in clinical research: Value of adverse events relatedness to study treatment: Analyses of data from randomized double-blind placebo-controlled clinical trials. *Ann Oncol*. 2017;28(6):1183–90.
49. Lignon J, Sibon D, Madelaine I, Brice P, Franchi P, Briere J, et al. Rituximab, dexamethasone, cytarabine, and oxaliplatin (R-DHAX) is an effective and safe salvage regimen in relapsed/refractory B-cell non-hodgkin lymphoma. *Clin Lymphoma, Myeloma Leuk*. 2010;10(4):262–9.
50. Locke FL, Miklos DB, Jacobson CA, Perales MA, Kersten MJ, Oluwole OO, et al. Axicabtagene Ciloleucel as second-line therapy for large B-cell lymphoma. *N Engl J Med*. 2022;386(7):640–54.
51. Lühring M, Schäfer I. Die Differenzierung von Nebenwirkungen im Rahmen klinischer Prüfungen. *Pharm Medizin*. 2014;3(Oktober):146–54.
52. Macedo AF, Marques FB, Ribeiro CF. Can decisional algorithms replace global introspection in the individual causality assessment of spontaneously reported ADRs? *Drug Saf*. 2006;29(8):697–702.
53. Macedo AF, Marques FB, Ribeiro CF, Teixeira F. Causality assessment of adverse drug reactions: Comparison of the results obtained from published decisional algorithms and from the evaluations of an expert panel. *Pharmacoepidemiol Drug Saf*. 2005;14(12):885–90.
54. Mounier N, El Gnaoui T, Tilly H, Canioni D, Sebban C, Casasnovas RO, et al. Rituximab plus gemcitabine and oxaliplatin in patients with refractory/relapsed diffuse large B-cell lymphoma who are not candidates for high-dose therapy. A phase II lymphoma study Association trial. *Haematologica*. 2013;98(11):1726–31.

55. Mouton JP, Mehta U, Rossiter DP, Maartens G CK. Interrater agreement of two adverse drug reaction causality assessment methods: A randomised comparison of the Liverpool Adverse Drug Reaction Causality Assessment Tool and the World Health Organization-Uppsala Monitoring Centre system. *PLoS One*. 2017;12(2).
56. Mukherjee SD, Coombes ME, Levine M, Cosby J, Kowaleski B, Arnold A. A qualitative study evaluating causality attribution for serious adverse events during early phase oncology clinical trials. Vol. 29, *Investigational New Drugs*. 2011. p. 1013–20.
57. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther*. 1981;30(2):239–45.
58. Paul-Ehrlich-Institut (PEI). Bericht Über Unerwünschte Ereignisse [Internet]. [cited 2024 Jan 11]. Available from: [https://www.pei.de/SharedDocs/Downloads/DE/arsneimittelsicherheit/pharmakovigilanz/meldebogen-cov2mab.pdf?\\_\\_blob=publicationFile&v=6](https://www.pei.de/SharedDocs/Downloads/DE/arsneimittelsicherheit/pharmakovigilanz/meldebogen-cov2mab.pdf?__blob=publicationFile&v=6)
59. Philip T, Guglielmi C, Hagenbeek A, Somers R, Van Der Lelie H, Bron D, et al. Autologous bone marrow transplantation as compared with salvage chemotherapy in relapses of chemotherapy-sensitive non-hodgkin's lymphoma. *N Engl J Med*. 1995;333(23):1540–5.
60. Russmann S, Kullak-Ublick GA, Weiler S, Stoller R, Egbring M. *Arzneimitteltherapie*. 2., überar. Schödl U, editor. Stuttgart: Thieme; 2016.
61. Sehn LH. Paramount prognostic factors that guide therapeutic strategies in diffuse large B-cell lymphoma. *Hematology Am Soc Hematol Educ Program*. 2012;2012:402–9.
62. Sharma S, Gupta AK, Reddy GJ. Inter-rater and intra-rater agreement in causality assessment of adverse drug reactions: a comparative study of WHO-UMC versus Naranjo scale. *Int J Res Med Sci*. 2017;5(10):4389–894.
63. Théophile H, André M, Miremont-Salamé G, Arimone Y, Bégaud B. Comparison of three methods (an updated logistic probabilistic method, the Naranjo and Liverpool algorithms) for the evaluation of routine pharmacovigilance case reports using consensual expert judgement as reference. *Drug Saf*. 2013;36(10):1033–44.
64. U.S. Food and Drug Administration. Good pharmacovigilance practices and pharmacoepidemiologic assessment [Internet]. Guidance for industry. 2005 [cited 2022 Aug 17]. Available from: <https://www.fda.gov/media/71546/download>
65. Wang E, Gulbis A, Hart JW, Nieto Y. The emerging role of gemcitabine in conditioning regimens for hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2014;20(9):1382–9.

66. Wiholm BE. The Swedish drug-event assessment methods. Special workshop-regulatory. *Drug Inf J.* 1984;18:267–9.
67. World Health Organization-The Uppsala Monitoring Center. The use of the WHO-UMC system for standardised case causality assessment [Internet]. 2009 [cited 2022 Aug 15]. p. 1–3. Available from: [https://who-umc.org/media/164200/who-umc-causality-assessment\\_new-logo.pdf](https://who-umc.org/media/164200/who-umc-causality-assessment_new-logo.pdf)

## 8 Publikationen

Im Rahmen des Promotionsvorhabens arbeitete der Autor dieser Dissertation mit der Abteilung Pharmakovigilanz der Niveau Studie zusammen und führte die Zweitbewertungen und Einordnungen der schwerwiegenden unerwünschten Ereignisse durch. Dies wurde unter anderem für die Evaluation des bis dahin bekannten Toxizitätsprofils der Therapie genutzt und Ergebnisse in der folgenden Publikation veröffentlicht:

Roch Houot, Viola Poeschel, Bettina Altmann, Stephanie Angel, Lorenz Thurner, Thomas Illmer, Marc Andre, Martin Dreyling, Hervé Maisonneuve, Hervé Tilly, Stephanie Mayer, Olivier Casasnovas, Steven Le Gouill, Fritz Offner, Guillaume Cartron, Andrea Kerkhoff, Thomas Weber, Joerg Hoffmann, Marita Ziepert, Wolfram Klapper, Emmanuel Itti, Dirk Hellwig, **Giorgi Natchkebia**, Laurence de Leval, Andreas Rosenwald, Corinne Haioun, Laurent Dercle, Philippe Gaulard, Gerhard Held Prolonged Remissions After Nivolumab Plus Gemcitabine/Oxaliplatin in Relapsed/Refractory T-cell Lymphoma. *HemaSphere*. 2022;6(2):E672. doi:10.1097/HS9.0000000000000672

## 9 Anhang

### 9.1 Beurteilungsbogen aus der NIVEAU Studie

#### SAE Assessment

Protocol (short title):	NIVEAU [DSHNHL 2015-1]		
Eudra-CT:	2016-002272-27		
Centre-ID			
Patient-ID			
SAE-No.			
CASE-ID			

<b>SAE Identification</b>			
<input type="checkbox"/> Initial – report	<input type="checkbox"/> Follow-up – report	Date of onset of event:  DD MM YYYY	
<b>SAE Assessment</b>			
Arm: <input type="checkbox"/> Standard <input type="checkbox"/> Experimentell	Serious?  <input type="checkbox"/> yes <input type="checkbox"/> no	Causal relationship? 1 = certain; 2 = likely; 3 = possible; 4 = unlikely; 5 = not related  ┌	Expected for treatment with Nivolumab?  <input type="checkbox"/> yes <input type="checkbox"/> no
Safety issue that might alter the current benefit-risk assessment?		<input type="checkbox"/> no <input type="checkbox"/> yes	
Safety issue that is sufficient to consider changes in the conduct of the trial?		<input type="checkbox"/> no <input type="checkbox"/> yes	
<input type="checkbox"/> further information will be required: Please specify:			
<b>SAE Assessment (Narrative)</b>			
Date (DD/MM/YYYY)	Name (Capitals or Stamp)		Signature (Resp. Person for 2 <sup>nd</sup> Assessment)



<b>NIVEAU</b> [DSHNHL 2015-1] Prof. G. Held, Homburg (DSHNHL) Sponsor: Saarland University, Saarbrücken, Germany		<b>SERIOUS ADVERSE EVENT</b>		<b>SAE</b> page 2 of 3			
Country-ID Hospital-ID:    -   -   -   -   - NIVEAU    NIV   -   -   -   -   - Pat-ID		SAE-Nº:   -   -   -		EudraCT-Nº: 2016-002272-27			
<b>5. Relevant medical history</b> (if possible, provide diagnosis)		<b>Onset Date</b> (dd/mm/yyyy)	<b>End Date</b> (dd/mm/yyyy)	<b>Continuing?</b>			
I.				<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown			
II.				<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown			
III.				<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown			
IV.				<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown			
V.				<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown			
<b>6. Concomitant therapy</b> (no SAE treatment)  (relevant medication, which is important in connection with occurrence and progression of the SAE, but was not given to treat SAE)		<b>Date and time of application</b> (if ongoing, record "cont" in column „End Date“)		<b>Route of administration</b>	<b>Single dose</b>	<b>Unit</b>	<b>Frequency</b> (e.g. 1x/d)
		<b>Onset Date</b> (dd/mm/yyyy)	<b>End Date</b> (dd/mm/yyyy)				
I.							
Indication:							
II.							
Indication:							
III.							
Indication:							
IV.							
Indication:							
V.							
Indication:							
VI.							
Indication:							
VII.							
Indication:							
VIII.							
Indication:							
+49 <span style="background-color: black; color: black;">                    </span>							



## 10 Dankausgang

An dieser Stelle möchte ich mich ganz herzlich allen beteiligten Personen meinen großen Dank aussprechen, die die Entstehung dieser Dissertationsarbeit ermöglicht haben.

An erster Stelle gilt mein besonderer Dank meinem Doktorvater Herrn Prof. Dr. Held für seine engagierte Betreuung, die wertvollen Ratschläge und seine enorme Unterstützung bei der Umsetzung der gesamten Arbeit. Seine Faszination mit Medizin und Wissenschaft wirkte immer ansteckend auf mich.

Ein großer Dank geht an alle Mitarbeiter der NIVEAU Studie für die angenehme und unkomplizierte Zusammenarbeit. Es war mir eine Ehre, ein Teil dieses Teams gewesen zu sein.

Mein ganz besonderer Dank gilt meinen Eltern für Ihre selbstlose Unterstützung in jeder Phase meines Lebens. Sie dienten als Vorbild und prägten meine Persönlichkeit. Alles was ich habe und alles was ich bin habe ich meinen Eltern zu bedanken.

Diese Arbeit widme ich aber meiner Frau, Ana. Denn eins ist klar: diese Arbeit wäre ohne sie nicht entstanden. Danke für den bedingungslosen Beistand und Vertrauen in mich in jeder Lebenslage.

## **11 Lebenslauf**

Aus datenschutzrechtlichen Gründen wird der Lebenslauf in der elektronischen Fassung der Dissertation nicht veröffentlicht.

Tag der Promotion: 08.04.2024

Dekan: Univ.-Prof. Dr. med. Michael D. Menger

Berichterstatter: Prof. Dr. Gerhard Held

Prof. Dr. Stefan Wagenpfeil

Prof. Dr. Markus Hecht