

Statistical Analysis Plan

- Final Version -

**Wirksamkeit und Sicherheit von Akupunktur bei
gonarthrosebedingten chronischen Schmerzen:
Multizentrische, randomisierte, kontrollierte Studie**

30. August 2004

**Wirksamkeit und Sicherheit von Akupunktur bei
gonarthrosebedingten chronischen Schmerzen:
Multizentrische, randomisierte, kontrollierte Studie**

We, the undersigned, have read the Statistical Analysis Plan and agree that it contains all information necessary for the statistical analysis of the trial.

Datum, Unterschrift

PD Dr. U. Mansmann

Dipl.-Math. S. Witte

Prof. Dr. N. Victor

Inhaltsverzeichnis

1	ZIEL DES ANALYSEPLANS.....	4
2	VERANTWORTLICHE FÜR DIE AUSWERTUNG.....	4
3	DETAILS DER STATISTISCHEN ANALYSE	5
3.1	DATENGRUNDLAGE FÜR DIE ANALYSE.....	5
3.2	DEFINITIONEN	5
3.3	ANALYSEDATENSÄTZE (ANALYSIS SETS).....	5
3.4	ENDPUNKTE.....	6
3.5	VERWENDETE SOFTWARE	8
4	DATA HANDLING.....	8
5	STATISTIK	8
5.1	INFERENZSTATISTIK (PRIMÄRE, FAS, PP, QS).....	8
5.2	SENSITIVITÄTSANALYEN (FAS, PP, QS).....	10
5.3	EXPLORATIVE STATISTIK	11
5.4	TABELLEN, LISTEN, ABBILDUNGEN	12
5.5	TABELLEN	13
5.6	LISTEN.....	13
5.7	ABBILDUNGEN.....	13
6	APPENDIX: TABELLEN SPEZIFIKATIONEN (HANDSCHRIFTLICH)	13

Abkürzungen

Abkürzung	Definition
CRF	Case Report Form
DSMB	Data safety and monitoring board
FAS	Full analysis set
IMBI	Institut für Medizinische Biometrie und Informatik
PP	Per protocol
QS	Quality set, Qualitätspopulation
SAE = SUE	Serious adverse event, schwerwiegendes unerwünschtes Ereignis
SAF	Safety set,

1 Ziel des Analyseplans

Einige Annahmen, Methoden und Prozeduren für die Statistische Auswertung sind bereits im Studienprotokoll beschrieben. Dieser Analyseplan spezifiziert die zu verwendenden statistischen und biometrischen Vorgehensweisen / Prozeduren im Detail.

2 Verantwortliche für die Auswertung

- a) Verantwortlichkeiten: Die Statistische Analyse wird vom IMBI eigenverantwortlich von den Biometrikern U. Mansmann und S. Witte durchgeführt. U. Mansmann ist der verantwortliche Biometriker. Es ist die Aufgabe der Biometriker, die Qualität der im Analyseplan aufgeführten statistischen Methoden und deren Umsetzung zu verantworten.

- b) Kontakt: (Institut für Med. Biometrie und Informatik, INF 305, 69120 Heidelberg)
 - PD Dr. rer. nat. Ulrich Mansmann
email: mansmann@imbi.uni-heidelberg.de
Tel: 06221/ 56-4128

 - Dipl.-Math. Steffen Witte
email: witte@imbi.uni-heidelberg.de
Tel: 06221/ 56-4371

 - Prof. Dr. rer. nat. Norbert Victor
email: victor@imbi.uni-heidelberg.de
Tel: 06221/ 56-4141

3 Details der statistischen Analyse

3.1 Datengrundlage für die Analyse

Die statistische Analyse wird durchgeführt, wenn der letzte Patient die Studie vollständig beendet hat und die Datenbank gemäß des Validierungsplanes bereinigt wurde. In die Analyse fließen Daten vom CRF und von den Telefoninterviews ein.

3.2 Definitionen

Begriff	Definition
Studienknie (study related knee)	Das für den WOMAC Fragebogen ausgewählte Knie.
Behandeltes Knie	Knie, dass mittels Akupunktur behandelt wird (mindestens ein obligat gestochener Punkt)
Related Knee	Kellgren 2-3, ACR für > 6 Monate erfüllt

3.3 Analysedatensätze (analysis sets)

Für die Analysen werden vier verschiedene Analysedatensätze definiert: Safety Set (SAF), Full Analysis Set (FAS), Per-Protocol Set (PP), Quality Set (QS).

- Safety Set (SAF) enthält alle Patienten, die in die Studie eingeschlossen werden (unterschiedene Einverständniserklärung).
- Full Analysis Set (FAS) enthält alle Patienten die randomisiert wurden und eine Akupunktur (Verum oder Sham) bzw. ein Behandlungsgespräch (Standard) erhalten haben. Patienten, die direkt nach der Randomisierung ihr Einverständnis zurückgezogen haben, gehören nicht in diesen Datensatz.
- Quality Set (QS) enthält alle randomisierten Patienten, aus sogenannten Qualitätszentren. Ein Zentrum wird als Qualitätszentrum bezeichnet, wenn
 - Mindestens 5 randomisierte Patienten (Rationale: Erfahrung)
 - Keine Entblindung gemäß Telefoninterview (Rationale: verantwortungsvoller Umgang mit der Studie)
 - Alle Akupunkturpatienten wurden jeweils 20-40 Minuten mit 7-15 Nadeln (Verum, ein Knie), 14-26 Nadeln (Verum, beide Knie) oder 9-10 Nadeln (Sham) behandelt
 - Kein Therapiewechsler im Zentrum
- Per-Protocol (PP) Set enthält alle randomisierten Patienten, die protokollgemäß behandelt wurden und die Interviews zur Erfassung der Zielgrößen abgeschlossen haben. Ausgeschlossen werden Patienten aus der Per-Protokoll-Population, wenn mindestens eines der folgenden Kriterien erfüllt ist:
 - Ein- oder Ausschlusskriterien verletzt.
 - Der explizite Wunsch des Patienten, die Therapie zu beenden.
 - Der explizite Wunsch des Patienten, die Therapie zu wechseln (insbesondere auch wenn ein Patient in der Standardgruppe trotzdem Akupunktur erhalten hat).
 - Das Nichteinhalten der Mindestanzahl von Behandlungssitzungen

- Anwendung explizit ausgeschlossener Behandlungen (Einnahme von Corticoiden, Einnahme von sonstigen Analgetika außer Diclofenac und Rofecoxib, Injektionen jeder Art im Bereich beider Beine, Moxa, Schröpfen und Elektroakupunktur)
- Dosisüberschreitung in der Standardtherapie (bis zu 150 mg/die Diclofenac bei Bedarf, alternativ Rofecoxib 25 mg/die, die Dauer ist nicht eingeschränkt).
- Der Patient verweigert weitere Telefoninterviews.
- Der Patient ist telefonisch nach wiederholten Versuchen nicht erreichbar, so dass das letzte Interview nicht durchgeführt werden kann.
- Der Prüfarzt verstößt gegen das vorgegebene Akupunkturschema (Dauer: 20-40 Minuten, Verum: 7-15 Nadel bei einseitiger Behandlung, 14-26 Nadeln bei zweiseitiger Behandlung, Sham: 9-10 Nadeln)
- Patient wurde nach eigenen Angaben (siehe Abschlusstelefoninterview) vom Prüfarzt bezüglich der Akupunkturtherapie entblindet.

3.4 Endpunkte

Folgende Endpunkte werden in der statistischen Analyse berücksichtigt:

Primärer Endpunkt:

Patienten, die einen der folgenden Gründe erfüllen, gelten als Therapieversager:

- Eine Verbesserung von Baseline bis zum 26-Wocheninterview um 36% bezüglich des WOMAC kann *nicht* erreicht werden.
- Therapiewechsel (unabhängig vom Grund)
- Nichtvorliegen des WOMAC aus der 26. Woche (drop out)
- Anwendung explizit ausgeschlossener Behandlungen (Einnahme von Corticoiden, Einnahme von sonstigen Analgetika außer Diclofenac und Rofecoxib, Injektionen jeder Art im Bereich beider Beine, Moxa, Schröpfen und Elektroakupunktur) zur Therapie von gonarthrosebedingten Beschwerden
- Einnahme von Diclofenac, Rofecoxib oder eine andere Schmerztherapie (z.B. Akupunktur) in den letzten drei Wochen vor Erfassung des Zielkriteriums in der 26. Woche
- zusätzliche Durchführung einer Akupunkturtherapie außerhalb der beschriebenen Behandlungsschemata
- Dosisüberschreitung: mehr als 150 mg/die Diclofenac bzw. 25mg/die Rofecoxib
- Speziell bei Akupunktur-Patienten: Gesamtdosis von 1000 mg Diclofenac bzw. 175 mg Rofecoxib im Zeitraum 3. Woche bis 3 Wochen vor Erfassung des Zielkriteriums wird überschritten (i.d.R. 3.-23. Woche). Werden beide Wirkstoffe in diesem Zeitraum eingenommen wird das Kriterium ($\text{Diclofenac [mg]} + 6 \cdot \text{Rofecoxib [mg]} > 1000 \text{ [mg]}$) angewandt (Diese Regelung schließt die Notfallmedikation mit ein).

Sekundäre Endpunkte:

- 1 Erfolgsrate basierend auf der Veränderung des WOMAC vom Ausgangswert zum Wert nach Beginn der 13. Woche (36%ige Verbesserung gilt als Erfolg)
- 2 Veränderung des WOMAC vom Ausgangswert zum Wert nach Beginn der 13. bzw. 26. Woche
- 3 Erfolgsrate basierend auf dem GPA nach Beginn der 7., 13. bzw. 26. Woche nach Behandlungsbeginn; Vergleich zwischen den Behandlungsgruppen (1, 2 oder 3 gilt als Erfolg auf der 6-Punkte-Skala)

- 4 SF12 (Lebensqualität): Veränderung in den Befindlichkeitsprofilen und im Gesamtscore vom Ausgangswert zum Wert nach Beginn der 13. und 26. Woche. Fehlen Teilfragen des SF12, so ist der Fragebogen nicht auswertbar und er gilt als nicht vorliegend.
- 5 Von Korff Chronic Pain Score (Von Korff, Ormel et al. 1992): Dieses Instrument besteht aus 7 Fragen zum Messen der Intensität von chronischem Schmerz und zum Messen der dadurch implizierten Beeinträchtigung. Neben der Anzahl der Schmerztage werden 4 weitere Kenngrößen analysiert (Analysezeitpunkte: CPI – Baseline, 3-Monate, 6-Monate).
 - 5.1 Anzahl der Schmerztage (0-92) (4. Frage)
 - 5.2 characteristic pain intensity = CPI (0-100): $10 * (\text{Mittelwert aus den ersten 3 Fragen})$
 - 5.3 disability score = DS (0-100): $10 * (\text{Mittelwert aus den letzten 3 Fragen})$
 - 5.4 disability points = DP (0-6): umkodiertes Ergebnis der 4. Frage und der letzten 3 Fragen
Umkodierung des disability scores (0,1,2,3):
[0-30) : 0 Punkte, [30,50) : 1 Punkt, [50, 70) : 2 Punkte, [70, 100]: 3 Punkte
Umkodierung der 4. Frage (0,1,2,3):
[0,7) : 0 Punkte, [7,15) : 2 Punkte, [15,31) : 3 Punkte, [31,184]: 4 Punkte
Die Antwort der 4. Frage bezieht sich auf 3 Monate, im von Korff bezieht sich diese Frage auf 6 Monate, daher müssen die Ergebnisse der 4. Frage vor der Umkodierung verdoppelt werden.
 - 5.5 Gesamtklassifikation, classification (Grad 0 – Grad IV):

Grad 0	:	CPI = 0,	DP = 0
Grad I	:	CPI < 50,	DP < 3
Grad II	:	CPI >= 50,	DP < 3
Grad III	:		DP >= 3
Grad IV	:		DP >= 5

Beispiel F1=1, F2=2, F3=3, F4=40, F5=5, F6=6, F7=7,
characteristic pain intensity = 20, disability score = 60,
disability points = 5, classification = 4
- 6 Vergleich der Responder und Non-Responder (i.S.v. WOMAC-Verschlechterung und/oder i.S.v. Hauptzielkriterium) anhand der Ausgangskenngrößen (Geschlecht, Alter, Dauer, WOMAC, ...) (2 Tabellen)
- 7 Anzahl der AU-Tage
- 8 Analgetikaverbrauch
- 9 Anzahl der Arztbesuche (zusätzlich zu den studienbedingten Arztbesuchen)
- 10 Anzahl unerwünschter Ereignisse
- 11 Anzahl schwerwiegender unerwünschter Ereignisse

Folgende Kenngrößen dienen der Beschreibung der chinesischen Diagnose für alle Patienten:

- TCM-Syndrom-Diagnose des Studienknies
- Kreuztabelle TCM-Syndrom-Diagnose und Kellgren-Score (2 oder 3) für das Studienknie mit 4 zweiseitigen 95% Konfidenzintervallen für die Wahrscheinlichkeit einen Kellgren von 2 zu erhalten.
- Kreuztabelle der Häufigkeiten von „Verschlechterung“ und „Schmerzlinderung“ des Studienknies.
- Nierenschwäche

Folgende Kenngrößen dienen der Beschreibung der Verumakupunktur:

- Durchschnittliche Dauer der Akupunktursitzungen (Verum- und Shamakupunktur)
- Anzahl der Fernpunkte der ersten Sitzung (nur Verumakupunktur)

- Anzahl der Ah-Shi-Punkte der ersten Sitzung (nur Verumakupunktur)

3.5 Verwendete Software

Alle Auswertungen werden mit dem Softwarepaket SAS[®] System 8.2 (SAS Inc., Cary/NC, USA) durchgeführt.

4 Data Handling

Fehlende Werte

Generell gilt: Fehlende Informationen werden *nicht* ersetzt, der entsprechende Wert wird auf „missing“ gesetzt. Beim WOMAC müssen pro Kategorie mindestens 80% der Fragen beantwortet sein (d.h. maximal 1 fehlender Wert von 5 Fragen in der ersten Kategorie, 0 fehlende Werte in der zweiten Kategorie und maximal 3 fehlende Werte von 17 Fragen in der dritten Kategorie), sonst gilt der WOMAC als nicht vorliegend.

Ausreißer

Im Rahmen des Querymanagements wird versucht, Extremwerte zu identifizieren. Der Prüfarzt muss ggf. Korrekturen vornehmen. Laborwerte werden in dieser Studie nicht analysiert, so dass entsprechende Messfehler nicht besonders berücksichtigt werden müssen.

5 Statistik

Die statistischen Methoden dienen der Darstellung der Rekrutierung, Datenqualität, Homogenität der Behandlungsgruppen, sowie der Darstellung der Wirksamkeit und Sicherheit der Behandlungen. Die meisten Auswertungen werden nach Datenbankschluss auf der Basis des „full analysis sets“ nach „Intention to Treat“ Prinzipien durchgeführt. Im Rahmen der Inferenzstatistik wird original SAS output ausgegeben, um dadurch die Transparenz zu erhöhen.

5.1 Inferenzstatistik (primäre, FAS, PP, QS)

Die Analyse des primären, binären Endpunktes basiert auf dem „full analysis set“. PP und QS werden als Sensitivitätsanalysen verwendet (nicht konfirmatorisch).

- Alle Tests hinsichtlich des Therapieunterschieds werden auf dem statistischen Niveau von 5% durchgeführt.
- Die jeweilige Testung der Paarvergleiche wird nach dem Abschlusstestprinzip vorgenommen.
- Damit garantiert die Testprozedur den Gesamtfehler erster Art von 5%.
- Im statistischen Modell werden folgende Stratifizierungsvariablen berücksichtigt: Zentrum und Anzahl behandelte Knie.

Bei sehr niedrigen Rekrutierungszahlen in einzelnen Zentren (weniger als 20 Patienten) werden Patienten zu größeren Einheiten zusammengefasst (mind. 20 Patienten). Hierzu wird ein nearest-neighbourhood-Ansatz zum Bilden von *Regionen* verwendet: Basierend auf den ersten beiden Ziffern der Postleitzahlen werden Zentren zu regionalen Clustern zusammengefasst: (10), (12), (13,14), (20,21,22,23,24), (26,27,29), (28), (30), (31,32), (33), (34,35,59), (40), (41), (42,51), (44), (45), (46,47), (48,49), (50,53), (55,56), (58), (60,61,63,64,65), (67,68), (69), (70), (71,72,74,75,76), (80,81), (82,83,85), (90), (91,92), (96).

Die regionalen Clustern enthalten mindestens 20 randomisierte Patienten und werden in der Hauptanalyse zur Adjustierung verwendet.

Folgende Global-Hypothese und die dazugehörigen Paarvergleichshypothesen werden formuliert:

$H_0^{VA-SA-ST}$: Die Erfolgsraten der Verum-Akupunktur (VA), der Sham-Akupunktur (SA) und der Standardtherapie (ST) sind gleich.

H_0^{VA-SA} : Die Erfolgsrate der Verum-Akupunktur (VA) ist gleich der Erfolgsrate unter Sham-Akupunktur (SA).

H_0^{VA-ST} : Die Erfolgsrate der Verum-Akupunktur (VA) ist gleich der Erfolgsrate unter der Standardbehandlung (ST).

H_0^{SA-ST} : Die Erfolgsrate der Sham-Akupunktur (SA) ist gleich der Erfolgsrate unter der Standardbehandlung (ST).

Ablauf der Testprozedur:

1. Globaltest für WOMAC
2. Falls Globaltest für WOMAC ...
 - a. ... nicht signifikant ausfällt, findet kein weiterer Test statt (Schlussfolgerung: keine bedeutsamen Unterschiede in der Wirksamkeit)
 - b. ... signifikant ausfällt, werden die Paarvergleichshypothesen zu WOMAC getestet (Spezifizierung der gefundenen Unterschiede)

Mit der Analyse werden die individuellen (nicht die marginalen) Veränderungen modelliert und somit die individuellen Odds Ratios (OR) interpretiert.

- Globaltest: Um Stratifizierung und Kovariablen berücksichtigen zu können, wird eine logistische Regression gerechnet. Es wird der Likelihood-Ratio-Test (2 Freiheitsgrade) zwischen dem Modell ohne Therapievariable (zusätzlich zu den oben genannten Variablen) und dem Modell mit den drei Therapieformen (zusätzlich zu den oben genannten Variablen) verwendet. Im Falle dessen signifikanten Ausganges folgen drei Paarvergleiche zwischen den drei Therapieformen:
- VA-SA: Analog zum Globaltest wird eine logistische Regression durchgeführt mit den entsprechenden Adjustierungsvariablen. Der Test ist wiederum ein Likelihood-Ratio-Test zwischen dem Modell ohne Therapievariable und dem mit einer binären Therapievariablen (1 für VA Patienten, 0 für SA Patienten). Im Falle eines signifikanten Testausganges wird H_0^{VA-SA} abgelehnt. Ein Vorteil der VA gegenüber der SA gilt als belegt, falls der Koeffizient zum Therapieeffekt positiv ist. Die Unterlegenheit der VA gegenüber der SA gilt als belegt, falls der Koeffizient zum Therapieeffekt negativ ist.
- VA-ST: Analog zum Globaltest wird eine logistische Regression durchgeführt mit den entsprechenden Adjustierungsvariablen. Der Test ist wiederum ein Likelihood-Ratio-Test zwischen dem Modell ohne Therapievariable und dem mit einer binären Therapievariablen (1 für VA Patienten, 0 für ST Patienten). Im Falle eines

signifikanten Testausganges wird H_0^{VA-ST} abgelehnt. Es wird von einem Vorteil der VA gegenüber der ST gesprochen, falls der Koeffizient zum Therapieeffekt positiv ist. Es wird von einer Unterlegenheit der VA gegenüber der ST gesprochen, falls der Koeffizient zum Therapieeffekt negativ ist.

- SA-ST: Analog zum Globaltest wird eine logistische Regression durchgeführt mit den entsprechenden Adjustierungsvariablen. Der Test ist wiederum ein Likelihood-Ratio-Test zwischen dem Modell ohne Therapievariable und dem mit einer binären Therapievariablen (1 für SA Patienten, 0 für ST Patienten). Im Falle eines

signifikanten Testausganges wird H_0^{SA-ST} abgelehnt. Es wird von einem Vorteil der SA gegenüber der ST gesprochen, falls der Koeffizient zum Therapieeffekt positiv ist. Es wird von einer Unterlegenheit der SA gegenüber der ST gesprochen, falls der Koeffizient zum Therapieeffekt negativ ist.

Die konditionale Modellierung in SAS wird mit dem Macro %glimmix vorgenommen:

```
%glimmix(data = efficacy, procopt=method=ML, stmts = %str(
  class treatment_group center;
  model womac_success = treatment_group bothknees / solution cl;
  random center;
),error = binomial, link = logit);
```

Für den Fall der Notwendigkeit für eine Interpretation der Ergebnisse im Sinne einer Nicht-Unterlegenheit, wird ein Konfidenzintervallverfahren verwendet. Um den "Points to consider on switching between superiority and non-inferiority" (PtC 1999, CPMP/EWP/482/99) gerecht zu werden, wird a priori eine Nicht-Unterlegenheitsgrenze für den Gruppenunterschied des WOMAC festgelegt. Eine Veränderung um mindestens 8 Punkten auf der WOMAC-Skala (0-100) wird als klinische relevant angesehen. Für die Übertragung auf die Odds-Ratio-Skala wurde ein parametrischer Bootstrap angewandt. Die notwendigen Annahmen zur Normalverteilung (Erwartungswert(baseline) = 48, Erwartungswert(6Monate, Standardgruppe) = 40, Varianz = 400, Kovarianz = 255) des WOMAC sind der Publikation Angst et al. (Angst F, Aeschlimann A, Stucki G, Arthritis and Rheumatism 2001; 45:384-391) entnommen. Aus dem Bootstrap ergibt sich eine Nichtunterlegenheitsschranke von $\delta_{clin}=0.4$. D.h. die Hypothese lautet $H_0: OR(Akupunktur/Standard) \leq 0.40$. Aufgrund von Ergebnissen in der Literatur ist die indirekte Bestätigung der Wirksamkeit möglich. Zieht man placebokontrollierte Studien mit Diclofenac als historischen Vergleich mit der Annahme eines konstanten Effektes heran (McKenna et al., Scand J Rheumatol 2001; 30(1):11-18), kann mit einer Grenze von 8.8 für den Unterschied des WOMAC auf die Wirksamkeit geschlossen werden. Die Übertragung mittels Bootstrap auf die OR-Skala liefert $\delta_{P(0)}=0.36$. Somit kann indirekt auf die Wirksamkeit der Akupunktur schließen zu können.

5.2 Sensitivitätsanalysen (FAS, PP, QS)

Zur Untersuchung der Stabilität der Inferenzstatistik werden einige Sensitivitätsanalysen durchgeführt (3x5=15 Analysen).

- 1 ML-Schätzung, und Modellvergleich über LQ-Test, logistische Regression mit dem random-effect wie in der Hauptanalyse
 - model womac_success = treatment_group bothknees baselinewomac
 - model womac_success = treatment_group bothknees baselinewomac treatment_group*baselinewomac
- 2 ML-Schätzung mit Standard logistischer Regression:

- ```
proc genmod data = efficacy descending;
 class treatment_group center;
 model womac_success = treatment_group bothknees center /
 dist = binomial;
run;
```
- ```
proc genmod data = efficacy descending;  
  class treatment_group center;  
  model womac_success = treatment_group bothknees center  
  treatment_group*center / dist = binomial;  
run;
```

3 Marginale Modellierung, GEE:

- ```
proc genmod data = efficacy descending;
 class treatment_group center;
 model womac_success = treatment_group bothknees / dist = binomial;
 repeated subject = center / type = cs;
run;
```

## 5.3 Explorative Statistik

### Sekundäre Analyse (FAS und PP):

Alle Zusatzauswertung (Sekundäre Endpunkte) haben nur explorativen Charakter. Folgende Auswertungen sind vorgesehen.

- Der WOMAC wird univariat zu den verschiedenen Zeitpunkten sowohl als dichotomisierte Zielgröße (wie Hauptzielgröße) als auch als ordinale Zielgröße ausgewertet.
- Die beiden durch den SF12 ermittelten Dimensionen werden graphisch dargestellt. Dies erlaubt einen Vergleich zwischen den Gruppen und eine Beobachtung der Veränderung über die Zeit.
- Der GPA wird univariat zu den verschiedenen Zeitpunkten sowohl als dichotomisierte Zielgröße (1-3 gilt als Erfolg) als auch als ordinale Zielgröße ausgewertet.
- AU-Tage: Ein Vergleich zwischen den Behandlungsgruppen wird mittels Abschlusstest durchgeführt: Kruskal-Wallis Test als Globaltest und Paarvergleiche durch Mann-Whitney-Test.
- Als Begleitvorhaben findet eine Analyse der Zusammenhänge der Konstrukte GPA (auch dichotomisiert), WOMAC (auch dichotomisiert), SF12 und von Korff Chronic Pain Score statt. Es werden (unter anderem) einfache Spearman'sche Korrelationskoeffizienten angegeben.
- Die 5 oben genannten Zielgrößen des von Korff Chronic Pain Score werden in den 3 Studienarmen verglichen (deskriptiv). Konfidenzintervalle der Mediane (oder bei entsprechender Verteilung der Mittelwerte) werden berechnet. Vergleichende Tests werden nur für die Gesamtklassifikation durchgeführt (unstratifiziert/unadjustiert: Kruskal-Wallis bzw. Wilcoxon-Mann-Whitney-Test für die Paarvergleiche).

### Sicherheitsanalyse (SAF/FAS)

Die Definition von unerwünschten Ereignissen ist im Protokoll gegeben. Alle UE von dem Unterzeichnen der Patienteneinverständniserklärung bis zur Abschlussuntersuchung (Final-Bogen) werden erfasst. UEs, die vor der ersten Behandlung eingetreten sind, werden nicht tabelliert (ICH-E3), daher wird in den Tabellen die FAS zu Grunde gelegt. Im Bericht wird die Anzahl dieser UEs explizit genannt, bei Bedarf beschrieben und im Listing gekennzeichnet.

Alle AEs werden nach Schweregrad, kausalem Zusammenhang und nach MedDRA-Organiklasse aufgelistet. Beschreibungen der Patienten mit schwerwiegenden unerwünschten Ereignissen werden erstellt (patient narratives).

Sicherheitsparameter für die drei Behandlungen sind:

- Häufigkeit der schwerwiegenden UE (SUE)
- Häufigkeit der unerwünschten Ereignisse (UE)
- Häufigkeit der UE nach der Einstufung des kausalen Zusammenhang gegliedert
- Häufigkeit der UE nach dem Schweregrad gegliedert
- Häufigkeit der UE nach dem Typ (MedDRA SOC) gegliedert

#### **5.4 Tabellen, Listen, Abbildungen**

Es wird ein Flow-Chart gemäß CONSORT-Statement erstellt. Die Originalwerte der Patienten werden in Listen dargestellt. Alle stetigen und ordinalen Variablen werden mindestens mit Stichprobenumfang, Mittelwert bzw. Median, Standardabweichung, Minimum und Maximum, die kategorialen Variablen werden mit Anzahl und Anteil (in %) in Tabellen getrennt nach den drei Behandlungsgruppen, dargestellt. Zur Verdeutlichung können Graphiken erzeugt werden.

##### Nominale und kategoriale Variablen

Alle nominalen und kategorialen Variablen werden zur Darstellung statistisch/deskriptiv zusammengefasst. Falls nicht anders spezifiziert, werden folgende Kenngrößen berechnet: Anzahl der Beobachtungen, absolute und relative Häufigkeiten basieren auf der Anzahl der Beobachtungen.

##### Stetige und Ordinale Variablen

Alle nominalen und kategorialen Variablen werden zur Darstellung statistisch/deskriptiv zusammengefasst. Falls nicht anders spezifiziert, werden folgende Kenngrößen berechnet: Anzahl der Beobachtungen, arithmetisches Mittel, Standardabweichung, Median, p5-, p25-, p75-, p95-Perzentile, Minimum und Maximum.

## 5.5 Tabellen

Die folgenden Tabellen werden für den Bericht erstellt. In den Fußnoten soll erkenntlich sein, ob die entsprechenden Daten aus dem CRF oder dem Telefoninterview stammen. Tabellenspezifikationen sind handschriftlich im Anhang.

## 5.6 Listen

Sämtliche Patientendaten werden in Listen dargestellt.

## 5.7 Abbildungen

Mindestens folgende Abbildungen werden zur Verdeutlichung erstellt.

1. Figure 1: Recruitment by time (SAF)
2. Figure 2: Flow-Chart gemäß CONSORT statement
3. Figure 3: Medianer Verlauf der Skalenwerte by group für WOMAC (FAS)
4. Figure 4: Medianer Verlauf der Skalenwerte by group für SF12 (FAS)
5. Figure 5: Medianer Verlauf der Skalenwerte by group für von Korff (FAS)
6. Figure 6: Medianer Verlauf der globalen Score by group für WOMAC (FAS)
7. Figure 7: Medianer Verlauf der globalen Score by group für SF12 (FAS)
8. Figure 8: Medianer Verlauf der globalen Score by group für von Korff (FAS)

## 6 Appendix: Tabellen Spezifikationen (handschriftlich)