



Standard Operating Procedure (SOP) für die Zusammenarbeit im Rahmen klinisch-wissenschaftlicher Studien

Rationale: Dieses „Standard Operating Procedure“ regelt die Zusammenarbeit zwischen dem Biostatistik Service des Forschungsmanagements und Technologietransfer (FMTT) der PMU und dem Anfragenden. Ziel ist die Festlegung eines Qualitätsstandards für die optimale Zusammenarbeit, Auswertung und Publikation.

- 1. Statistische Bearbeitung medizinischer Forschungsprojekte/Studien:** Das Angebot der statistischen Betreuung und Bearbeitung wissenschaftlicher Studien kann von allen Angehörigen der Kliniken und Institute der Paracelsus Medizinischen Privatuniversität (PMU) sowie von externen Wissenschaftlern in Anspruch genommen werden.
- 2. Erstbesprechung:** Ziel ist zunächst eine Klärung der Vorgehensweise. Diese kann je nach Fortschritt der Studie unterschiedlich sein: befindet sich die Studie z.B. in der Planungsphase, so werden Themen wie Ethikkommissionsantrag, saubere Datenerfassung mittels Datenbank, Planung der notwendigen Fallzahlen oder Studiendesign (offen, blind, randomisiert, kontrolliert, mono- oder multizentrisch, Pilotstudie etc.) angesprochen.

Sind die Daten bereits erhoben, werden die wesentlichen Fragestellungen und Ziele der Studie geklärt. Zum Erstberatungsgespräch sollten alle vorliegenden Projektunterlagen (z.B. Projektbeschreibung, Liste der zu untersuchenden Fragestellungen, Datenerhebungsbogen, etc.) mitgebracht werden. Das Erstgespräch ist kostenlos. Im Rahmen des Beratungsgesprächs wird über eine künftige Zusammenarbeit im Projekt entschieden. Der beratende Statistiker behält sich vor, bei gravierenden Mängeln aus statistischer Sicht die Zusammenarbeit abzulehnen.

- 3. Datenübernahme/Datenhandling:** Ein Schema für korrekte Daten sehen Sie auf Seite 5. Dies sollen die nachfolgenden Punkte anhand eines Beispiels veranschaulichen.
 1. Es werden ausschließlich anonymisierte oder pseudo-anonymisierte Daten übernommen. Insbesondere dürfen Vornamen, Nachnamen oder Geburtsdatum in den Daten nicht aufscheinen.
 2. Die Daten müssen vorher sorgfältig bereinigt werden (data cleaning). Dies ist entscheidend wichtig, damit die Datenauswertung überhaupt begonnen werden kann.
 3. Fehlerbehaftete Daten, wie z.B. Tippfehler, inkonsistente Daten etc. werden nicht akzeptiert und wieder zur neuerlichen Überprüfung zurückgeschickt.
 4. Überprüfen Sie Ihre Daten am besten mittels eines Excelfilters oder indem Sie die Daten selbst vorher deskriptiv auswerten und sich Minimum und Maximum der Werte ansehen. Fallen Ihnen zweifelhafte Werte auf, so entscheiden Sie ob Sie diese Werte als valide akzeptieren und somit in den Daten drinnen lassen oder diese entfernen (z.B. wenn die Messung nicht akzeptiert werden kann).
 5. Fehlende Werte: Lassen Sie die Felder für fehlende Werte einfach frei und verwenden Sie keine Codes wie n/a, 999 oder dgl. um diese zu markieren.
 6. Verwenden Sie als Variablenbezeichnung Bezeichnungen, die möglichst kurz und prägnant sind sowie – falls notwendig - den entsprechenden Einheiten (mmHg, mm, ml, etc.). Dies ist vor allem bei



der Erstellung der Abbildungen (Figures der Publikation), sodass die Beschriftungen gleich von Anfang an korrekt sind und nicht nachher noch umbezeichnet und nachformatiert werden müssen.

7. Wichtig: In der ersten Zeile müssen die Variablen durchnummeriert werden. Beachten Sie: es dürfen keine Lücken in der Durchnummerierung sein. Im Weiteren wird nun stets über die Variablennummern statt der Variablen gesprochen. Dies erleichtert und beschleunigt die Auswertung beträchtlich.
 8. Zu den Daten muss ein Tabellenblatt in Excel erstellt werden, das Bedeutung und Codierung der Variablen erklärt. Dies sollte möglichst kurz und prägnant sein.
 9. Formulieren Sie die Fragestellungen/Hypothesen nun mittels der Variablennummern. Diese werden ebenfalls in Excel als Tabellenblatt eingefügt. Nummerieren Sie die Fragestellungen durch, sodass später leichter auf diese Bezug genommen werden kann.
 10. Das nachfolgende Streichen oder eine Hinzunahme von weiteren Probanden zu den ursprünglichen Daten kann nicht akzeptiert werden.
 11. Zwecks Nachweis der Übereinstimmung der für die Analyse verwendeten Daten mit den Originaldaten, werden seitens des Biostatistik-Services alle übermittelten Datensätze im Originalzustand mit Datumslabel archiviert.
- 4. Auswertung, Darstellung und Publikation der Studie:** Die komplette Auswertung der Studiendaten wird mittels modernster mathematisch-statistischer Methoden und Spezialsoftware durchgeführt (STATISTICA, StatXact, NCSS, PASS, MATHEMATICA, SPSS, R, ...).
- 5.** Die Abbildungen der Ergebnisse werden für die Publikation in medizinischen Fachjournalen in 600dpi Druckqualität und optimal für den Submissionsprozess vorbereitet.
- 6.** Die Ergebnisse der Auswertung werden schriftlich per E-Mail als Bericht in Form von Excel oder Worddateien übermittelt.
- 7. Ziel ist die Publikation in medizinischen Fachjournalen oder die Approbation Ihrer Dissertation/Habilitation/Professur.**



Wir ersuchen Sie den nachfolgenden Appendix „Dokumentation der statistischen Beratung“ auszufüllen, auszudrucken und unterzeichnet per E-Mail an wolfgang.hitzl@pmu.ac.at zu schicken.

Appendix

SOP Nr:

Dokumentation der statistischen Beratung

Bitte beachten Sie, dass mit der Auswertung erst begonnen werden kann, sofern das SOP vollständig und korrekt ausgefüllt, unterzeichnet und zurück geschickt wurde.

Datum des Erstgesprächs: _____

Name des Rechnungsempfänger, bzw. Firma mit Rechtsform: _____

Rechnungsanschrift: _____

Tel.-Nr. und E-Mail Adresse: _____

Falls Dissertation/Diplomarbeit:

Name des/der Dissertant*in/Diplomand*in _____

E-Mailadresse für Rechnungszustellung: _____

- 1.** Die Rechnung wird über eine SALK-Kostenstelle beglichen: nein: , ja: ,
falls ja, bitte SALK-Kostenstellenummer angeben: _____
- 2.** Die Rechnung wird über eine PMU-Kostenstelle beglichen: nein: , ja: ,
falls ja, bitte PMU-Kostenstellenummer angeben: _____

3. **Kurzer Arbeitstitel der Studie:**

- 4.** **Stadium der Arbeit:** Planungsphase; Datenauswertung;
Votum der Ethikkommission liegt vor: Ja Nein; Nachbearbeitung; Anderes

- 5.** **Verwendung der Arbeit:** Diplomarbeit Dissertation Vortrag/Kongress Habilitation
 Einreichung eines Antrags bei der Ethikkommission Publikation Anderes

6. **Welche Beratung erwarten Sie als PI?**

1. Datenerfassung: Programmierung einer ACCESS-Datenbank
2. Erstellung einer Randomisierungsliste für die Studienteilnehmer
3. Unterstützung bei einer Diplomarbeit/Dissertation
4. Studienplanung: Beratung bei der Wahl des Studiendesigns



5. Fallzahlplanung: Berechnung der optimalen Fallzahlen bei vorgegebener Power der Studie und klinisch relevantem Unterschied
6. Ergänzung des statistischen Teils von Ethikkommissionsanträgen
7. Durchführung der statistischen Datenverarbeitung, diese umfasst:
- a. Kontrolle der Daten auf Ausreißer
 - b. Auswahl von moderner statistischer Auswertungsmethoden und Modellen
 - c. Dokumentation der Ergebnisse
 - d. Erstellung druckfertiger Graphiken für die Publikation
 - e. Beschreibung der statistischen Methoden (z.B. für das zu publizierende Paper)
 - f. Korrekte Interpretation der Ergebnisse aus mathematisch-statistischer Sicht
 - g. Kritische Revision des Papers
 - h. Finale Freigabe des Manuskripts zur Publikation
 - i. Verteidigung des Papers gegenüber den Reviews („comments to the reviewer“)
8. Verteidigung eines papers gegenüber Editor/Reviewer
9. Sonstiges

In Übereinstimmung mit der „Richtlinie zur Sicherung guter wissenschaftlicher Praxis an der PMU“ wird bestätigt, dass im Fall der Durchführung der statistischen Datenverarbeitung (Punkte 7 und/oder 8) der verantwortliche Statistiker als Koautor in allen Veröffentlichungen geführt wird, in welchen die Ergebnisse verwendet werden. Wird die Publikation ohne Nennung des Autors veröffentlicht, erfolgt eine Meldung gemäß „Richtlinie zur Sicherung guter wissenschaftlicher Praxis“ an den Dekan für Forschungsangelegenheiten. Dies wird zusammen mit einer Begründung dem PI und dem Klinik- bzw. Institutsleiter schriftlich mitgeteilt.

Kostenvoranschlag:

Beschreibung zu Art und Umfang der nötigen Arbeiten, diese umfassen: Erstgespräch, Besprechungen, Datenauswertungen, Erstellung von Figures, kritisches Lesen und Freigabe des Manuskripts, Verteidigung gegenüber Reviewers

Sorgfältig vorbereitete und saubere Daten, klare Fragestellungen und eindeutige Zuordnungen der Variablen zu den Fragestellungen erlauben eine wesentlich schnellere Bearbeitung der Analysen, Darstellungen und erhöht die Effizienz und Effektivität.

Wir ersuchen Sie höflich dies sorgfältig zu beachten!

Erfüllungszeitraum: Beginn der Auswertung spätestens 14 Tage nach Erhalt des korrekt formatierten und fehlerfrei vorliegenden Datensatzes und Fragestellungen (siehe Punkt 3) und der elektronischen Übermittlung des unterzeichneten SOPs.

Datum:

Unterschrift des Principal Investigators:

Unterschrift des Biostatistikers:



Vorlage für Datenerfassung und Datenmatrix

1. **Durchgehende Variablenbezeichnung:** im Folgenden wird stets auf die Nummer statt der Variablen referiert, z.B. 4 = date of surgery. Dies schafft Klarheit in der Diskussion, Eindeutigkeit und so werden die Daten schnell und effizient ausgewertet!

Die Variablennamen bleiben selbstverständlich erhalten. Hängen Sie am Ende der Variablenbezeichnung (z.B. age), die jeweilige Variablennummer an, (z.B. 2) sodass am Ende dann steht: age_2!

Anmerkung: Sie können die Nummern leicht mit dem Befehl „Verketten“ in Excel anhängen.

Die Durchnummerierung **muss exakt passen**, sonst kann die Datenauswertung nicht durchgeführt werden und vor allem Deadlines für Ihre Studie nicht eingehalten werden, da es dann zu Verzögerungen kommt.

2. Verwenden Sie **Bezeichnungen, die eindeutig, kurz und prägnant** sind, möglichst mit passenden Einheiten, z.B. blood pressure (systolic) mmHg postop 1d

ID_1	age_2	gender_3	group_4	date of surgery_5	location_6	blood pressure (systolic) mmHg preop_7	blood pressure (systolic) mmHg postop 1d_8	time till death (days)_9	death (censor)_10
1	35.4	f	control	43018	temporal	122	128	120	0
2	65.8	m	treatment		inferior	142		363	0
3	78.7	f	treatment	42705	superior	132	135		1

3. bei **fehlenden Werten** lassen Sie die **Zelle frei**, geben Sie **nichts ein**, also **nicht** z.B. „n/a“, 999, oder dgl.
4. Verwenden Sie eindeutige Begriffe, d.h. diese müssen **stets gleich** sein, also z.B. nicht einmal inferior, dann „inf“ oder „infer.“ oder dgl.
5. Überprüfen Sie generell die Werte auf **Plausibilität** (siehe Kommentar auf Seite 1, Punkt 3.4).

Vorlage für die Fragestellungen Ihrer Arbeit

1. In den weitaus meisten Fällen werden mehrere Gruppen von Probanden miteinander verglichen, z.B. möchte man eine Gruppe „group“ (treatment vs control oder dgl) bzgl. ,blood pressure (pain, eye pressure, etc) im Verlauf vergleichen. In diesem Fall geben Sie bitte die Fragestellung wie folgt **kurz und prägnant** mithilfe der **Variablennummern** an.

Z.B. „7 und 8 soll bzgl. Gruppe 6 verglichen werden“, d.h. also „blood pressure (systolic) mmHg preop (7) und 1d postop im Verlauf (8) bzgl. der Gruppe „group“ (6). So können die Fragestellungen **viel schneller und effizienter fixiert werden.**

2. Eine andere typische Fragestellung sind Zusammenhänge/Unterschiede, z.B. gibt es Unterschiede bzgl. „location“ in beiden Gruppen. Einfach ausgedrückt heißt das: „5 bzgl. Gruppe 6“.
3. Oder es geht um Korrelationen, z.B. Gibt es eine Korrelation von „age“ und „blood pressure preop“ und „blood pressure post“, d.h. „Korrelation zwischen 2 7 und 8“.

Diese Vorlagen haben sich als im Laufe der Jahre als sehr effizient herausgestellt und führen dazu, dass Ihre Studie rasch und effizient ausgewertet werden kann.

Bei etwaigen Fragen stehe ich Ihnen gerne zur Verfügung:

PD Dr. rer.nat. Wolfgang Hitzl, MSc, MSc

Leiter of Team Biostatistik und Publikation von klinischen Studien//machine learning
Forschungsmanagement und Technologietransfer (FMTT)

E-Mail: wolfgang.hitzl@pmu.ac.at / Tel: 0699 144 200 32