

Institut für Klinische Epidemiologie und Biometrie
Am Schwarzenberg 15 / A15, 97078 Würzburg

Prof. Dr. rer. nat. Dr. rer. med.
Götz Gelbrich
Professor für Biometrie

Ethikkommission der
Medizinischen Fakultät der
Universität Würzburg

Tel.: 0931 / 201-46409
Fax: 0931 / 201-647310
Gelbrich_G@ukw.de
www.epidemiologie.uni-wuerzburg.de

Würzburg, 15.04.2021

Hinweise zu Fallzahlbegründungen bei Anträgen an die Ethikkommission

Positionspapier der Ethikkommission der Medizinischen Fakultät der Universität Würzburg

Bei der Beurteilung der Vertretbarkeit klinischer oder epidemiologischer Forschung am Menschen ist abzuwägen zwischen dem **Zweck** und dem voraussichtlichen **Nutzen** der Studie einerseits und der Art der medizinischen **Maßnahmen** und **Datenerhebungen** an den Studienteilnehmern sowie der **Zahl** der teilnehmenden Personen andererseits.

Aus **ethischer** Sicht ist eine **zahlenmäßig zu kleine Studie** unvertretbar, da die Studienfragestellung mit hoher Wahrscheinlichkeit nicht hinreichend genau beantwortet werden kann. Andererseits ist eine **zahlenmäßig zu große Studie** ebenfalls unvertretbar, weil die Studienfragestellung mit einer gewissen Teilnehmerzahl hinreichend genau beantwortet ist und die darüber hinaus einbezogenen Personen unnötig dem Risiko und den Belastungen ausgesetzt werden, die mit den medizinischen Maßnahmen und der Datenerhebung verbunden sind.

Aus **datenschutzrechtlicher** Sicht ist die Erhebung und Verarbeitung personenbezogener Daten nur dann gerechtfertigt, wenn damit ein Zweck erfüllt werden kann, für den die Datenerhebung geeignet ist. Kann mit einer zu kleinen Datenmenge die Studienfragestellung mit hoher Wahrscheinlichkeit nicht befriedigend beantwortet werden, so erfüllt die Datenerhebung den Zweck nicht und ist damit nicht gerechtfertigt. Lässt sich mit einer gewissen Datenmenge die Studienfragestellung hinreichend genau beantworten und damit der Zweck erfüllen, so ist eine darüberhinausgehende Datenerhebung durch den angegebenen Zweck nicht mehr zu rechtfertigen.

Aus diesen Gründen sind Anträge für **jede prospektive Forschung am Menschen** mit einer begründeten Angabe der **Fallzahl** zu versehen, die im **Studienprotokoll** enthalten sein sollte; andernfalls ist ein Antrag für die Ethikkommission nicht beurteilbar.

Die Begründung enthält Angaben

- zur **statistischen Genauigkeit**, mit der eine Fragestellung beantwortet werden soll (z.B. Breite eines Konfidenzintervalls, Prävalenzen von Eigenschaften in verschiedenen Gruppen, Unterschied zwischen Mittelwerten und Standardabweichung bei metrischen Merkmalen, Inzidenzen von Ereignissen in verschiedenen Gruppen); dies können realistisch erwartete Daten sein oder aber hypothetische Effekte, die von der Größe her klinisch relevant und daher es wert sind, überzeugend dargestellt zu werden

- zur **statistischen Zuverlässigkeit** (z.B. Konfidenzniveaus, Signifikanzniveau und Power, Likelihood-Ratios für verschiedene Datenszenarien)

Dabei ist es (insbesondere in nicht-konfirmatorischen Studien) nicht erforderlich, den klassischen Konventionen (Test einer Nullhypothese auf dem 5%-Signifikanzniveau mit mindestens 80% Power) zu folgen. Weder muss dem Schema des Signifikanztests gefolgt werden, noch ist beim Signifikanztest das 5%-Niveau zwingend. Es muss die statistische Aussagekraft der beabsichtigten Fallzahl dargestellt sein, damit die ethische Rechtfertigung durch das Studienziel beurteilt werden kann.

Auch bei Studien zur explorativen Forschung oder Pilotstudien wie auch bei der Datenerhebung an einer nach Personen feststehenden klinischen Kohorte sollte eine Fallzahlplanung erfolgen.

FALLZAHLPLANUNG NICHT MÖGLICH? – DOCH!

Häufig wird in Anträgen bei der Ethikkommission ausgeführt, dass im konkreten Fall eine Fallzahlbetrachtung nicht erforderlich oder nicht möglich sei. Das stimmt in der Regel nicht. Die folgenden Beispiele sollen zeigen, dass auch in solchen Fällen Fallzahlbetrachtungen sehr genau ausgeführt werden können. Diese Ausführungen stellen keinen Leitfaden oder Lehrbuchtext dar und beanspruchen keine Vollständigkeit. Wir gehen auch davon aus, dass nicht jeder Kliniker die Ausführungen bis ins Detail verstehen kann. Zweck der Beispiele ist aufzuzeigen, dass eine Fallzahlbetrachtung in der Regel möglich ist und dass es dafür sehr konkrete Vorstellungen gibt. Für die Klärung im Detail **suchen Sie bitte die biometrische Konsultation**. Fallzahlplanung gehört in den handwerklichen Bereich der Biometriker, nicht der Kliniker. (Das schließt nicht aus, dass Kliniker sich auch dieses Handwerkszeug aneignen; es ist jedoch nicht der Regelfall.)

BIOMETRISCHE UNTERSTÜTZUNG ZUR FALLZAHLPLANUNG

Alle Universitäten mit einer Medizinischen Fakultät verfügen über ein Institut, an dem das Fach Biometrie vertreten ist. Dort kann man professionelle Beratung erhalten. An der **Universität Würzburg** ist dies das Institut für Klinische Epidemiologie und Biometrie (IKEB, Direktor: Prof. Dr. Peter U. Heuschmann).

Antragsteller sind frei in der Wahl der Institution, deren Beratung sie in Anspruch nehmen wollen. Natürlich bietet es sich an, die Hilfe im eigenen Haus zu suchen, aber auch bei Partnern, mit denen bereits eine wissenschaftliche Kooperation besteht. Für die Bewertung bei der Ethikkommission ist es unerheblich, ob die biometrische Expertise an der Universität Würzburg oder an einer anderen Universität in Deutschland oder in Europa (beispielsweise multizentrische IITsd eingeholt wurde; entscheidend ist allein die inhaltliche Ausführung).

BEISPIELE

Explorative Studien. Hier ist die Fragestellung häufig, ob bestimmte Zusammenhänge überhaupt bestehen. Auch wenn es keinerlei Daten gibt, die Anhaltspunkte für die mögliche Stärke dieser Zusammenhänge geben, sind sie klinisch nur von Interesse, falls sie eine gewisse Mindeststärke besitzen. Diese Mindeststärke (Relevanzgrenze) kann Grundlage einer Fallzahlfestsetzung sein.

- **Beispiel:** Es wird gefragt, ob ein Biomarker für eine Diagnostik brauchbar sein könnte. Man möchte dazu die Werte des Markers bei Patienten und Kontrollen vergleichen. Man fordert dann Mindestwerte für Sensitivität und Spezifität, unterhalb derer eine diagnostische Brauchbarkeit nicht gegeben ist. Die Fallzahl setzt man so fest, dass eine Differenz der Mittelwerte von ($Z_{\text{Sens}} + Z_{\text{Spez}}$) Standardabweichungen gut erkannt werden kann (Z = Quantile der Standardnormalverteilung). Um die Entscheidung zwischen den Szenarien „brauchbar“ und „Marker hat gar nichts mit der Diagnose zu tun“ symmetrisch zu gestalten, könnte die Power auf 90% und das Signifikanzniveau auf einseitig 10% festgesetzt werden.
- **Beispiel:** Es wird gefragt, ob sich ein Biomarker im Patientenbefinden widerspiegelt. Als relevant könnte z.B. betrachtet werden, wenn mindestens 10% des Befindens durch den Marker erklärt werden. Man plant die Fallzahl daher für das signifikante Aufdecken einer Korrelation von $r=0,32$ zwischen Marker und einer geeigneten Lebensqualitätsskala.

Bei Signifikanztests in explorativen Studien ist eine Korrektur für multiples Testen in der Regel nicht erforderlich (Ausnahme: beim „gene hunting“ sinnvoll), während beim confirmatorischen Testen mehrerer Hypothesen das Signifikanzniveau in Abhängigkeit von der Testhierarchie zu korrigieren ist.

Pilotstudien. Sie finden statt, um zu entscheiden, ob und in welcher Weise die Forschung durch eine größere Studie fortgesetzt werden sollte. Diese Entscheidung wird an Kriterien festgemacht, die sich aus den Daten der Pilotstudie ergeben. Für eine sinnvolle Entscheidung müssen die Pilotdaten eine hinreichende Aussagekraft besitzen, die sich auch aus der Fallzahl ergibt. Folglich ist die Fallzahl zu planen, nur eben nicht unbedingt aufgrund der medizinischen Hauptfragestellung, sondern aufgrund der Fragestellung, die mit der Pilotstudie entschieden werden soll.

- **Beispiel:** Zur Vorbereitung eines größeren Protokolls stellt sich die Frage, ob die Patienten diesem Protokoll hinreichend folgen werden. Eine Pilotstudie kann dann den Zweck haben, zwischen den Szenarien „mindestens 80% der Patienten sind protokolltreu“ und „höchstens 50% der Patienten sind protokolltreu“ zu unterscheiden. In einer solchen Pilotstudie wäre Protokolltreue (ja/nein) der Endpunkt, und die Fallzahl sollte auf die Unterscheidung der Raten von 50% und 80% geplant sein. Der Endpunkt der späteren großen Studie spielt keine Rolle für die Fallzahl.
- **Beispiel:** In einer großen Studie sollen die Mittelwerte zweier Gruppen bezüglich eines quantitativen Endpunkts verglichen werden. Eine Fallzahlschätzung für eine große Studie ist aktuell nicht möglich, weil Daten zur Standardabweichung des Endpunkts fehlen. Deshalb wird eine Pilotstudie initiiert, um die Standardabweichung zu schätzen. Die Fallzahl der Pilotstudie wird so geplant, dass die Obergrenze der Standardabweichung mit einem bestimmten Konfidenzniveau (z.B. 95%) auf einen bestimmten Faktor genau (z.B. 1,2) geschätzt werden kann. – In diesem Fall ist es auch möglich, dass die „eigentliche“ Studie begonnen wird mit der Maßgabe, nach der Pilot-Fallzahl die Standardabweichung zu schätzen und auf dieser Basis per Protokollzusatz die Fallzahl der gesamten Studie zu kalkulieren.
- Oftmals steht die Frage, ob es sich lohnt, eine größere, confirmatorische Studie zu initiieren. Man muss sich also selbst vom Sinn einer solchen Studie überzeugen, nicht aber die wissenschaftliche Gemeinschaft von bestimmten Ergebnissen. Die Einhaltung des 5%-Signifikanzniveaus ist dann nicht erforderlich. Man könnte also die primäre medizinische Fragestellung vorab an einer kleineren Stichprobe z.B. zum einseitigen Signifikanzniveau von 20% mit einer Power von 80% prüfen, um für sich selbst den Glauben an die Szenarien „es gibt eher den vermuteten Effekt“ und „es gibt eher gar keinen Effekt“ zu quantifizieren. Alternativ könnten auch Likelihood-Methoden statt Signifikanztests herangezogen werden.

Bestehende Kohorten. Gern wird (oft im Rahmen medizinischer Dissertationen) das Langzeitergebnis einer Patientengruppe einer Klinik untersucht. Dafür werden die Patienten eines klinischen Kollektivs aus einem bestimmten Zeitraum kontaktiert und nachbefragt oder auch zu Nachuntersuchungen eingeladen. Hier stehen der potentiell teilnehmende Personenkreis und damit die Obergrenze der Fallzahl fest. Auch eine solche Untersuchung sollte wissenschaftliche Fragestellungen besitzen, und demnach lässt sich eine Poweranalyse durchführen, die zeigt, ob eine hinreichende Aussicht auf Beantwortung der Fragestellungen besteht.

Ist diese Aussicht nicht gegeben, so steht die ethische Rechtfertigung für die Inanspruchnahme der Patienten für studiengetriebene Untersuchungen und Datenerhebungen in Frage. Bei zu geringer Power sollte erwogen werden, eine multizentrische Auswertung unter Einbeziehung gleichartiger Kollektive aus anderen Kliniken durchzuführen.

Fallzahlfestsetzung aufgrund geübter Praxis. In bestimmten Situationen kann die Fallzahl aufgrund einer routinemäßigen Praxis festgesetzt werden. Auch dies muss aber begründet sein; die Routine allein genügt nicht.

- Beispiel: In klinischen Studien der Phase I wird zur Dosisfindung oft das „3+3 Design“ angewandt. Dieses ist in diesem speziellen Kontext etabliert und von Zulassungsbehörden akzeptiert. Die maximale Fallzahl ist dann schematisch durch die Zahl der untersuchten Dosisstufen bestimmt.
- Beispiel: Bei der Entwicklung von MRT-Sequenzen findet keine übliche statistische Auswertung statt. Es werden Algorithmen angepasst, bis gute Ergebnisse erzielt werden. Es hat sich gezeigt, dass üblicherweise 10-20 Testpersonen benötigt werden. Soweit einzig die MRT-Sequenzen entwickelt werden und keine klinische Fragestellung untersucht werden soll, wird für diesen Zweck die Angabe der routinemäßig festgesetzten Fallzahl akzeptiert.
- Der Verweis, dass es publizierte Arbeiten mit ähnlich gelagerten Fragestellungen und einer bestimmten Fallzahl gibt, ist zur Fallzahlbegründung nicht ausreichend, denn damit ist nicht belegt, dass die Fallzahl der publizierten Arbeit adäquat geplant war. Akzeptabel ist es aber, eine nachvollziehbare Fallzahlplanung aus der Literatur zu kopieren mit dem Argument, dass aufgrund der ähnlich gelagerten Fragestellung für die beantragte Studie dieselbe Effektgröße und dieselben statistischen Fehlergrößen zugrunde gelegt werden.

Existenzbeweis. Für den Nachweis, dass ein bestimmtes Phänomen überhaupt auftreten kann genügt es, einen Patienten mit diesem Phänomen zu beschreiben. Hierfür genügt also ein Fallbericht mit $n=1$.

Möchte man eine Patientenkohorte screenen, ob ein bestimmtes Phänomen (z.B. eine genetische Variante) überhaupt darin vorkommt, dann lautet die Frage: Wie viele Patienten muss man untersuchen, um mit einer bestimmten Sicherheit mindestens einen Patienten mit diesem Phänomen zu erwischen? Die Wahrscheinlichkeit, erfolgreich zu sein, hängt dabei von der Prävalenz des Phänomens in der Patientenpopulation ab. Man muss also angeben, bei welcher Prävalenz man ein Phänomen gern aufspüren würde.

Numerisches Beispiel: Wenn das Phänomen bei mindestens 5% der Patientenpopulation auftritt, möchte man eine Chance von 99% haben, dass mindestens einer der in der Studie untersuchten Patienten dieses Phänomen zeigt. Hierfür müsste man 90 Patienten untersuchen.

Deskriptive Studien. Im Bereich der patientennahen **Laborforschung** kann eine exakte Fallzahlbetrachtung schwierig sein, z.B., wenn die in Frage kommende Patientenzahl sehr gering ist, oder wenn die Auftrittshäufigkeit einer Marker-Expression unbekannt ist. Gleiches gilt für die Generierung von Tiermodellen aus humanen Tumorproben, oder sogar patienteneigenen PDX-Modellen (Patient-Tumor-Xenografts). In solchen Fällen sollte ausführlich dargelegt werden, aus welchen Gründen nur eine grobe Fallzahlschätzung möglich ist, und warum trotzdem mit einer sinnvollen und verwertbaren Aussage gerechnet werden kann. **Je geringer die Aussagekraft eines solchen Projektes ist, desto geringer darf auch die Belastung für den Patienten durch das Vorhaben sein.**

Laborforschungsprojekte ohne konkrete Fallzahlplanung sollten eine absolute Ausnahme bleiben, in der überwiegenden Mehrzahl der Projekte ist auch hier eine solide Kalkulation (z.B. auf Basis der vorliegenden Literatur/Publikationen) möglich.

VERZICHT AUF FALLZAHLBETRACHTUNG

- Soweit die Ethikkommission für ein Vorhaben eine **Kurzbescheinigung** ausstellt, dass ein Ethikvotum in diesem Fall nicht erforderlich ist und/oder wenn in dem Vorhaben keine Einwilligung betroffener Personen erforderlich ist oder die Einwilligung rein datenschutzrechtlicher Natur ist (Verarbeitung bereits vorhandener personenbezogener Daten), wird die Ethikkommission keine Fallzahlbetrachtung verlangen.

- Soweit die Belastung der Studienteilnehmer so geringfügig ist, dass jegliche Beeinträchtigung schutzwürdiger Interessen der Teilnehmer sicher auszuschließen ist, kann die Ethikkommission auch ohne Vorliegen einer Fallzahlbetrachtung die ethische Unbedenklichkeit bescheinigen. Dies stellt eine **Einzelfallentscheidung** dar. Antragsteller können daraus **keinen Anspruch** ableiten, dass in einem vermeintlich ähnlich gelagerten Fall auf die Vorlage einer Fallzahlplanung verzichtet werden müsse.
- Auch im Falle eines Verzichts auf die Vorlage einer Fallzahlbetrachtung durch die Ethikkommission wird Antragstellern **empfohlen**, die Durchführung einer solchen zu erwägen. Immer mehr Zeitschriften verlangen für eine **Publikation** Angaben zur Fallzahlplanung bzw. auch bei retrospektiven Studien mit limitierter Fallzahl eine Powerkalkulation.