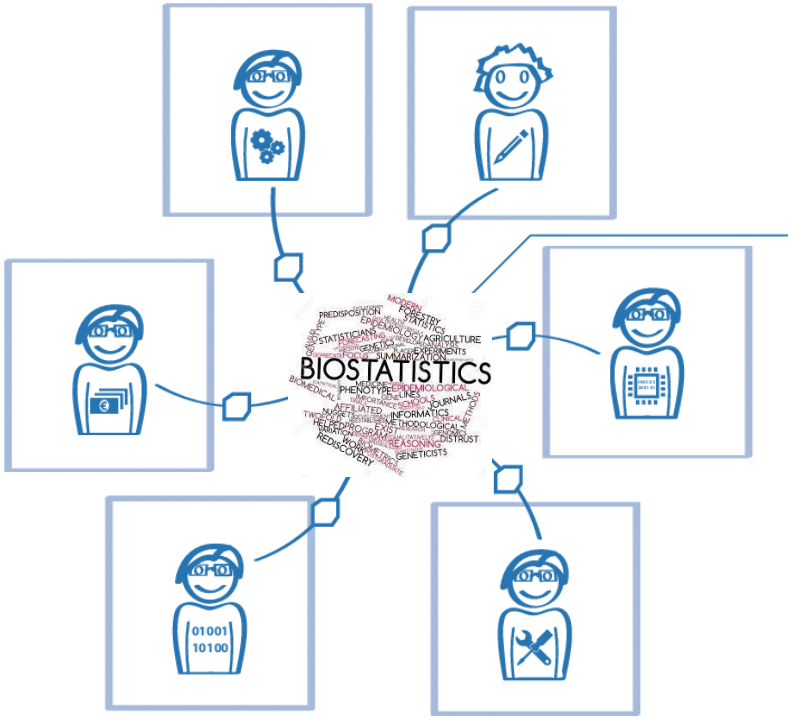
Компетенции специалиста по биостатистике — на стыке математики и медицины. Почему важен контекст при обработке биомедицинских данных?

Александр Солодовников, к.м.н.
Директор по управлению качеством ООО «Статэндокс»
Доцент кафедры профилактической и семейной медицины ФГБОУ ВО УГМУ МЗ РФ
Доцент кафедры биохимии ФГБОУ ВО РУДН

statandocs.com

Специалист по биомедицинской статистике – ожидания и требования индустрии

- Требования нормативно-правовых актов
- Дизайн клинических и наблюдательных медицинских исследований
- Обоснование объема выборки
- Статистическое планирование
- Обеспечение управления и целостности данных
- *Статистический анализ*
- *Визуализация результатов*



Пример

Нормативные правовые акты ЕАЭС

stat&docs

- Рекомендация Коллегии Евразийской экономической комиссии от 17 июля 2018 г. № 11 "О Руководстве по общим вопросам клинических исследований"
- Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 79 "Об утверждении Правил надлежащей клинической практики Евразийского экономического союза"
- Рекомендация Коллегии Евразийской экономической комиссии от 03 ноября 2020 г. № 19 "О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов"
- Рекомендация Коллегии Евразийской экономической комиссии от 12 марта 2019 г. № 8 "О Руководстве по подбору дозы лекарственных препаратов"
- Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 85 "Об утверждении Правил проведения исследований биоэквивалентности лекарственных препаратов в рамках Евразийского экономического союза"
- Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 89 "Об утверждении Правил проведения исследований биологических лекарственных средств Евразийского экономического союза"

Пример

Первичная конечная точка

- **Выбор** первичной переменной должен отражать ***принятые нормы и стандарты в соответствующей области медицины***.
- Рекомендуется использовать надежную и валидированную переменную, ***данные о которой получены в предыдущих исследованиях или опубликованы в научной литературе***.
- Необходимо представить достаточные доказательства того, что первичная переменная является **валидной и надежной мерой клинически значимой и важной пользы вмешательства** для ***популяции пациентов, описываемой с помощью критериев включения и невключения*** субъектов в исследование.
- **Первичную переменную следует использовать для определения размера выборки**, *если не обоснован иной подход.*

Рекомендация Коллегии Евразийской экономической комиссии от 03 ноября 2020 г. № 19 "О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов"

Пример

Первичная конечная точка

- Во многих случаях подход к оценке исходов может быть неочевидным и должен быть тщательно регламентирован (например, ***недопустимо без дополнительных уточняющих условий выбирать смертность в качестве первичной переменной***; смертность можно оценивать, *сравнивая доли живых в фиксированные временные точки или сопоставляя общие распределения времени выживания в течение установленного интервала*).
- Другим распространенным примером является **повторяющееся событие**; мерой эффекта вмешательства может быть ***простая дихотомическая переменная*** (любое возникновение события в течение установленного промежутка времени), ***время до первого возникновения события***, ***частота возникновения*** (число событий в единицу времени наблюдения) и т. д.

Рекомендация Коллегии Евразийской экономической комиссии от 03 ноября 2020 г. № 19 "О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов"

Пример

Первичная конечная точка

- ***Оценка функционального статуса*** субъекта исследования в динамике при изучении вмешательства ***при хронических заболеваниях*** вызывает затруднения при выборе первичной переменной.
- Существует **<u>много возможных подходов</u>**, таких как:
  - сравнения оценок, выполненных в начале и в конце интервала наблюдения;
  - сравнение для регрессионных прямых углов наклона, рассчитанных с помощью всех оценок в течение всего интервала наблюдения;
  - сравнения долей субъектов со значениями показателей, превышающими определенный порог или опускающимися ниже определенного порога,

    или
  - сравнения, основанные на методах для повторных измерений данных.

Рекомендация Коллегии Евразийской экономической комиссии от 03 ноября 2020 г. № 19 "О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов"

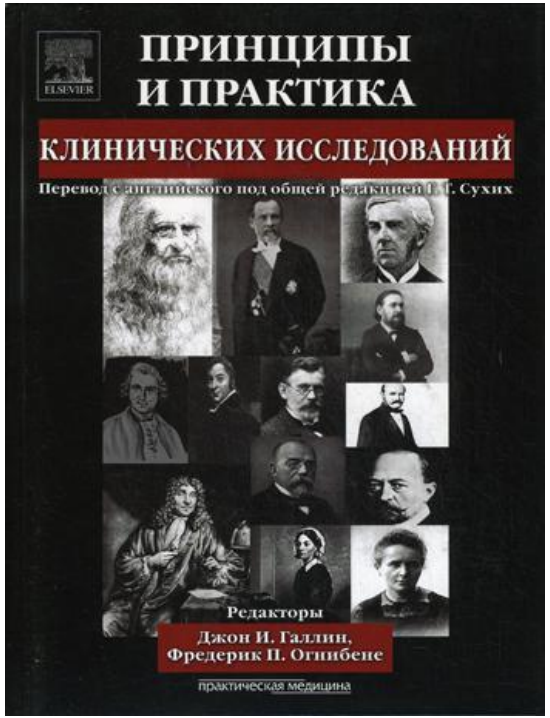
Пример

Первичная конечная точка

- В протоколе клинического исследования необходимо ***указать первичную переменную вместе с обоснованием ее выбора***.
- **Переформулировка первичной переменной** после *снятия ослепления почти всегда будет недопустима, поскольку привносимые вследствие этого систематические ошибки трудно поддаются оценке*.
- Если клинический эффект, определенный основной целью, ***будет измеряться несколькими методами***, в протоколе клинического исследования один **из методов измерений необходимо указать в качестве первичной переменной** *на основании клинической значимости, важности, объективности и (или) других применимых его характеристик – во всех случаях осуществимости такого выбора.*

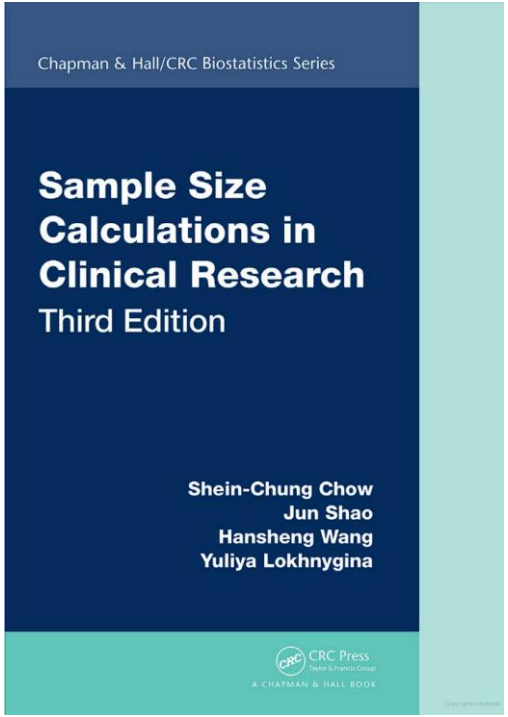
Рекомендация Коллегии Евразийской экономической комиссии от 03 ноября 2020 г. № 19 "О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов"

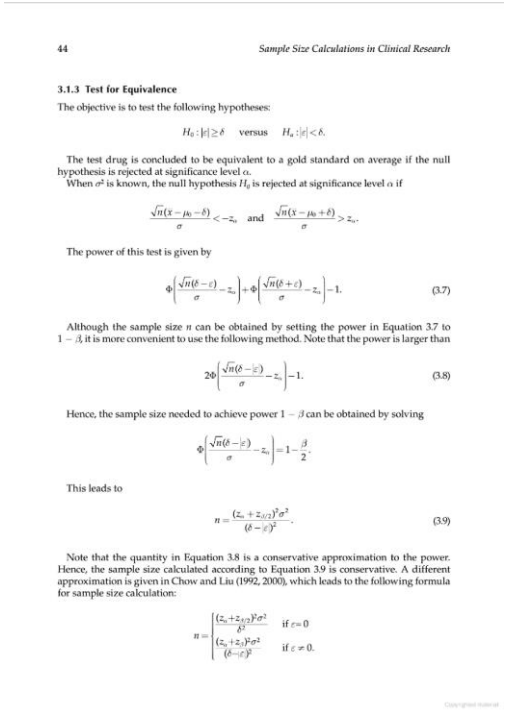
Дизайны медицинских исследований



Многообразие дизайнов

Обоснование и расчет размера выборки





44 *Sample Size Calculations in Clinical Research*

3.1.3 Test for Equivalence

The objective is to test the following hypotheses:

$$H_0: |\varepsilon| \geq \delta \quad \text{versus} \quad H_a: |\varepsilon| < \delta.$$

The test drug is concluded to be equivalent to a gold standard on average if the null hypothesis is rejected at significance level α.

When σ^2 is known, the null hypothesis H_0 is rejected at significance level α if

$$\frac{\sqrt{n}(\bar{x}-\mu_0-\delta)}{\sigma} < -z_\alpha \quad \text{and} \quad \frac{\sqrt{n}(\bar{x}-\mu_0+\delta)}{\sigma} > z_\alpha.$$

The power of this test is given by

$$\Phi\left(\frac{\sqrt{n}(\delta-\varepsilon)}{\sigma} - z_\alpha\right) + \Phi\left(\frac{\sqrt{n}(\delta+\varepsilon)}{\sigma} - z_\alpha\right) - 1. \tag{3.7}$$

Although the sample size n can be obtained by setting the power in Equation 3.7 to $1-\beta$, it is more convenient to use the following method. Note that the power is larger than

$$2\Phi\left(\frac{\sqrt{n}(\delta-|\varepsilon|)}{\sigma} - z_\alpha\right) - 1. \tag{3.8}$$

Hence, the sample size needed to achieve power $1-\beta$ can be obtained by solving

$$\Phi\left(\frac{\sqrt{n}(\delta-|\varepsilon|)}{\sigma} - z_\alpha\right) = 1 - \frac{\beta}{2}.$$

This leads to

$$n = \frac{(z_\alpha + z_{\beta/2})^2\sigma^2}{(\delta-|\varepsilon|)^2}. \tag{3.9}$$

Note that the quantity in Equation 3.8 is a conservative approximation to the power. Hence, the sample size calculated according to Equation 3.9 is conservative. A different approximation is given in Chow and Liu (1992, 2000), which leads to the following formula for sample size calculation:

$$n = \begin{cases} \dfrac{(z_\alpha + z_{\beta/2})^2\sigma^2}{\delta^2} & \text{if } \varepsilon = 0 \\ \dfrac{(z_\alpha + z_\beta)^2\sigma^2}{(\delta-|\varepsilon|)^2} & \text{if } \varepsilon \neq 0. \end{cases}$$

Пример

Обоснование и расчет размера выборки

stat&docs

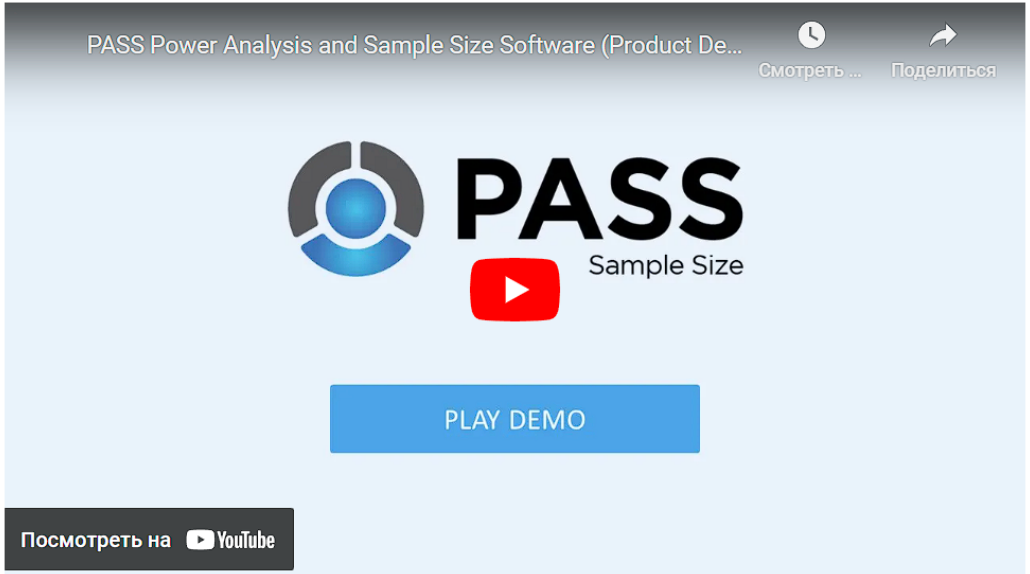
Sample Size & Power

PASS software provides sample size tools for over 1100 statistical test and confidence interval scenarios - more than **double the capability** of any other sample size software. Each tool has been **carefully validated** with published articles and/or texts.

Get to know PASS by downloading a free trial, viewing the video to the right, or exploring this website.

PASS comes complete with integrated documentation and PhD statistician support.

PASS has been fine-tuned for over 25 years, and is now the leading sample size software choice for clinical trial, pharmaceutical, and other medical research. It has also become a mainstay in all other fields where sample size calculation or evaluation is needed.



Статистическое планирование

"план статистического анализа" - документ, содержащий детализированное и описанное специализированным языком (с использованием технических терминов) изложение основных элементов анализа, указанного в протоколе, и включающий в себя детализированные процедуры проведения статистического анализа первичных и вторичных переменных, а также других данных;

117. План статистического анализа может быть представлен в виде отдельного документа, составляемого после завершения подготовки протокола клинического исследования. В плане статистического анализа можно осуществить более техническую и детализированную проработку основных характеристик, указанных в протоколе клинического исследования. План статистического анализа может содержать детализированные процедуры выполнения статистического анализа первичных и вторичных переменных и других данных. План статистического анализа необходимо пересматривать и обновлять по результатам ослепленной проверки данных (при необходимости) и окончательно оформить до снятия ослепления. Необходимо хранить документы о времени окончательного оформления плана статистического анализа, а также времени последующего снятия ослепления.

Рекомендация Коллегии Евразийской экономической комиссии от 03 ноября 2020 г. № 19 "О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов"

Пример

Статистическое планирование

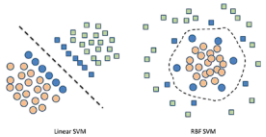
DETAILED STATISTICAL ANALYSIS PLAN (SAP)

1. Administrative information

1.1. Title, registration, versions and revisions

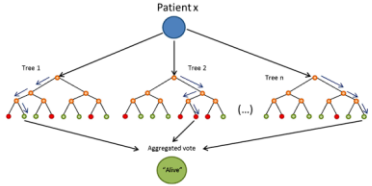
Full study title	A prospective observational study on the value of conventional hemodynamic parameters in estimating cardiac output and predicting mortality in critically ill patients
Acronym	Simple Intensive Care Studies-I (SICS-I)
Local project number	201500144
Clinicaltrials.gov number	NCT02912624
Study protocol version	1.1 (23 September 2016)
SAP version	1.0 (3 April 2018)
SAP revision history	None
SAP revision justification	-
SAP revision timing	Revision will be conducted after publishing the two main manuscripts. This revision will concern detailed SAPs for each sub-study.

Figure 2. Representation of a linear and a radial hyperplane in a Support Vector Machine classifier.



Class 1: squares, class 2: circles, and in blue the boundary cases.

Figure 3. Representation of a Random Forest classifier, with n trees



The circles represent the nodes, with the red and green circles signalling the terminal nodes, and the arrows show the "optimal split" computed for each node of each tree. The aggregated vote of all (n) trees is then combined and expressed as a final classification, for example, "alive".

Пример

Статистическое планирование

stat&docs

Primary analysis of primary endpoint:

The primary estimand is defined **to evaluate the hypothetical proper effect of XXX without additional effect of rescue therapy** (hypothetical strategy):

- Taking into account that patients would keep the benefit of the randomized treatment after **rescue medication**,
- taking into account that patients who are unable to continue taking the study drug **whatever the reason** would keep the benefit of the randomized treatment.

Thus, before performing ANCOVA in the FAS, a multiple imputation procedure as if patients would keep the benefit of the randomized treatment will be used to impute: missing data for the primary endpoint at 12 weeks (for patients with missing data at W8 and W12 and for patients with missing data at W12 only), and data collected after rescue therapy for non-controlled patients.

Пример

Статистическое планирование

stat&docs

Supplementary analysis of primary endpoint:

- This supplementary estimand is defined **to evaluate the effect of XXX**:
  - under the assumption that **rescue medication** was not introduced and assuming that patients will not continue to benefit from the treatment,
  - taking into account the unfavorable outcome when patients are unable to continue taking the study drug due to **lack of efficacy or AE**,
  - taking into account that patients who are unable to continue taking the study drug for **other reason** than lack of efficacy or AE would keep the benefit of the randomized treatment.

Before performing ANOVA in the FAS:

- a placebo-based multiple imputation (MI) approach will be used to impute values after rescue medication and for treatment discontinuation due to lack of efficacy or AE. The copy increment from reference approach will be applied: once the patient deviates, its post-deviation mean increments copy those from the placebo group. The patient's mean profile following deviation tracks that of the mean profile in the placebo group, but starting from the value of the criteria of interest already obtained ((the approach will be to impute values post rescue medication and post treatment discontinuation for AE or lack of efficacy by the slope of the placebo arm starting from the value of HbA1c observed at the withdrawal visit). Post deviation data in placebo group is imputed under randomized-arm MAR;

- a MI approach as if patients would keep the benefit of the randomized treatment will be used to impute values for treatment discontinuation for other reason.

Пример

Статистическое планирование

stat&docs

Two other additional analysis will be performed on the primary endpoint on FAS:

- This analysis is defined **to evaluate the effect of treatment policy of XXX + rescue medication VS placebo + rescue medication**
  - taken into account values collected after **rescue medication** (treatment policy),
  - taking into account that patients who are unable to continue taking the study drug **whatever the reason** would keep the benefit of the randomized treatment.

Thus, before performing ANCOVA, a multiple imputation procedure as if patients would keep the benefit of the randomized treatment will be used to impute missing data for the primary endpoint at 12 weeks. Here data collected after rescue therapy will be taken into account in the analysis.

- Last observation carried forward (LOCF) method will be used to impute post rescue evaluations and other missing data at 12 weeks for the primary endpoint.

The multiple imputation approach will be based on the regression method to impute values for each missing data at W8 and W12. Covariate use for the regression method will be auxiliary variables which might be predictive of both the chance of missing values and the underlying values themselves. Those additional auxiliary variables will be defined, before the blind is broken.

Управление данными (data management) и обеспечение целостности данных (data integrity)

- Рекомендация Коллегии Евразийской экономической комиссии от 19.09.2023 г. «О Руководстве по обеспечению целостности данных и валидации компьютеризированных систем»

8. Настоящее Руководство направлено на использование риск-ориентированного подхода к управлению данными, который включает риск, критичность и жизненный цикл данных. Пользователям настоящего Руководства необходимо понимать управление данными (как жизненный цикл), чтобы идентифицировать данные, оказывающие наибольшее влияние на процессы GMP. Исходя из этого, можно определить и внедрить наиболее эффективный и действенный контроль, основанный на оценке рисков, и обзор данных.

9. Рекомендации и методы, приведенные в данном руководстве, подготовлены на основе аналогичных руководств международных и региональных организаций (Системы сотрудничества фармацевтических инспекций (PIC/S), Всемирной организации здравоохранения (ВОЗ), Управления по контролю лекарственных средств и изделий медицинского назначения Великобритании (MHRA), Международного общества по фармацевтическому инжинирингу (ISPE), Ассоциации производителей лекарственных средств для парентерального применения (PDA)).

Управление данными (data management) и обеспечение целостности данных (data integrity)

Данные, на которых основаны любые решения, должны соответствовать критериям «ALCOA+», которые включают в себя:

прослеживаемость (A – attributable) данных до лица, создавшего запись;

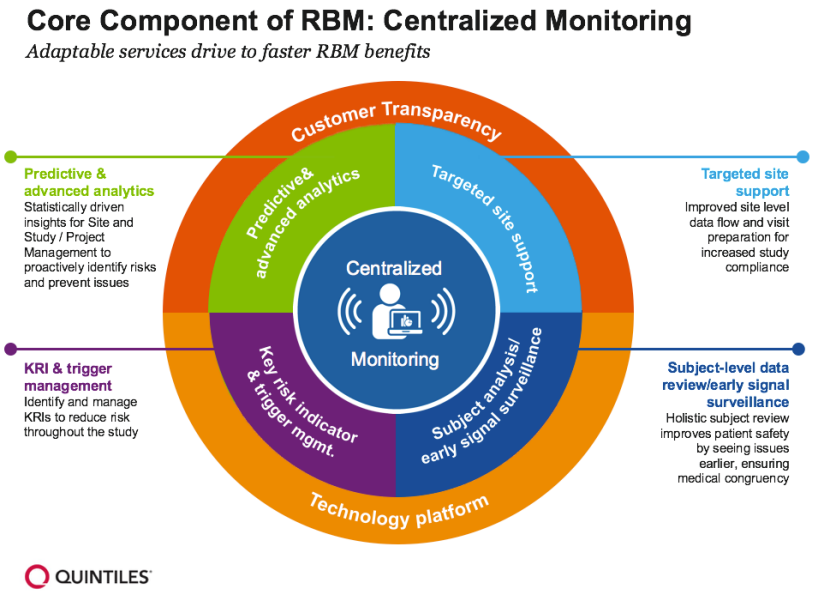
читаемость (L – legible) данных,

своевременность (C – contemporaneous) данных,

подлинность (O – original) данных

точность (A – accurate) данных;

соответствие дополнительным требованиям, обеспечивающим полноту (complete), последовательность (consistent), устойчивость (enduring) и доступность (available) данных (обозначенных в акрониме символом «+»).



Спасибо за внимание!

Александр Солодовников, директор по управлению качеством
e-mail: asolodovnikov@statandocs.com
mobile: +7 919 3776265
www: https://statandocs.com