

Почему многие опубликованные исследования неверны

Сергей Воронов

24 марта 2014 г.

Научные исследования и истинность соотношения

Исследование истинно	Соотношение верно	Да	Нет	Сумма
	Да		$(1 - \beta)R$	α
Нет		βR	$1 - \alpha$	$1 - \alpha + \beta R$
Сумма		R	1	$R + 1$

$$PPV = \frac{(1 - \beta)R}{(1 - \beta)R + \alpha}$$

$$R = \frac{\text{Количество экспериментов, когда соотношение было истинно}}{\text{Количество экспериментов, когда соотношение было ложно}}$$

Научные исследования и истинность соотношения (добавлено смещение)

Иссл. \ Соотн.	Да	Нет
Да	$(1 - \beta)R + u\beta R$	$\alpha + u(1 - \alpha)$
Нет	$(1 - u)\beta R$	$(1 - u)(1 - \alpha)$

$$PPV = \frac{(1 - \beta)R + u\beta R}{\alpha + u(1 - \alpha) + (1 - \beta)R + u\beta R}$$

$$R = \frac{\text{Количество экспериментов, когда соотношение было истинно}}{\text{Количество экспериментов, когда соотношение было ложно}}$$

u = Процент экспериментов, результаты которых не должны входить в результаты исследования, но, тем не менее, вошли.

Зависимость от размера выборки

Вывод

Чем меньше размер выборки, на которой производилось исследование, тем менее вероятно, что результаты верны.

Меньше выборка \Rightarrow меньше мощность \Rightarrow меньше PPV.

В некоторых областях не всегда возможно взять большую выборку для исследования (molecular predictors).

Зависимость от размера эффекта

Вывод

Чем меньше размер эффекта в научной области, в которой производилось исследование, тем менее вероятно, что результаты верны.

Меньше эффект \Rightarrow меньше мощность \Rightarrow меньше PPV.

В качестве примера:

- Влияние курения на некоторые формы рака и сердечно-сосудистых заболеваний (относительный риск 3-20)
- Влияние гена на некоторые мультигенные болезни (относительный риск 1.1-1.5)

Тенденция: все больше современных исследований в области генетики производится с относительным риском менее 1.05.

Зависимость от R

Вывод

Чем больше общее количество гипотез и меньше количество выбранных для проверки, тем менее вероятно, что результаты верны.

$$PPV = \frac{(1 - \beta)R}{(1 - \beta)R + \alpha}$$

Особо подвержены гипотезопорождающие исследования (например, предсказания с помощью микрочипов).

Зависимость от гибкости модели

Вывод

Чем более гибкие модели, определения, методы используются в исследовании, тем менее вероятно, что результаты верны.

Большая гибкость увеличивает вероятность того, что <отрицательный> результат станет <положительным> и наоборот \Rightarrow увеличивает u .

Производятся попытки стандартизировать инструменты для исследования; кроме того возможна манипуляция с публикацией части результатов или неверной интерпретации.

Зависимость от собственных интересов исследований

Вывод

Чем больше финансовая или иная заинтересованность и предубежденность в результатах, тем менее вероятно, что результаты исследования верны.

Конфликт интересов может увеличить смещение u .
Интересы не обязательно финансовые, а, например, принципиальная позиция в области некоторого исследования.

Зависимость от «популярности» области исследования

Вывод

Чем больше команд вовлечено в исследования по данной теме, тем менее вероятно, что результаты исследования верны.

Среди многих команд исследователей у одной из них может появиться выброс, в то время, как остальные не опубликуют результат, как незначимый.

Кроме того, некоторые команды могут поставить своей целью наибоьстрейшее представление «положительного» результата.

Значения PPV для различных модельных исследований

Table 4. PPV of Research Findings for Various Combinations of Power ($1 - \beta$), Ratio of True to Not-True Relationships (R), and Bias (u)

$1 - \beta$	R	u	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase V/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase V/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015

Предлагаемые меры

- Улучшение доказательной базы (т.е. большая выборка и уменьшение шума)
- Проверка широких гипотез
- Организация общей базы исследование (где стоит регистрировать как положительные, так и отрицательные результаты; возможно пререгистрация исследования)
- Разработка общего протокола исследований там, где это возможно
- Обращать внимание на R до исследования (подозрительно высокие R стоит проверить отдельно).

Ответ на статью: Goldman, 2007

$$\begin{aligned}
 \text{Байесовский фактор} &= \frac{\text{Вероятность истинности при отрицании } H_0}{\text{Вероятность истинности при условии } H_0} = \\
 &= \frac{\text{Pr(False neg)} \times \textit{bias} + \text{Pr(True pos)}}{\text{Pr(True neg)} \times \textit{bias} + \text{Pr(False pos)}} = \frac{\beta \times \textit{bias} + (1 - \beta)}{(1 - \alpha) \times \textit{bias} + \alpha}
 \end{aligned}$$