



Конфаундинг

Аглиуллина Саида Тахировна
доцент кафедры эпидемиологии и доказательной
медицины ФГБОУ ВО Казанский ГМУ
Минздрава России, к.м.н.

saida.agliullina@kazan-gmu.ru





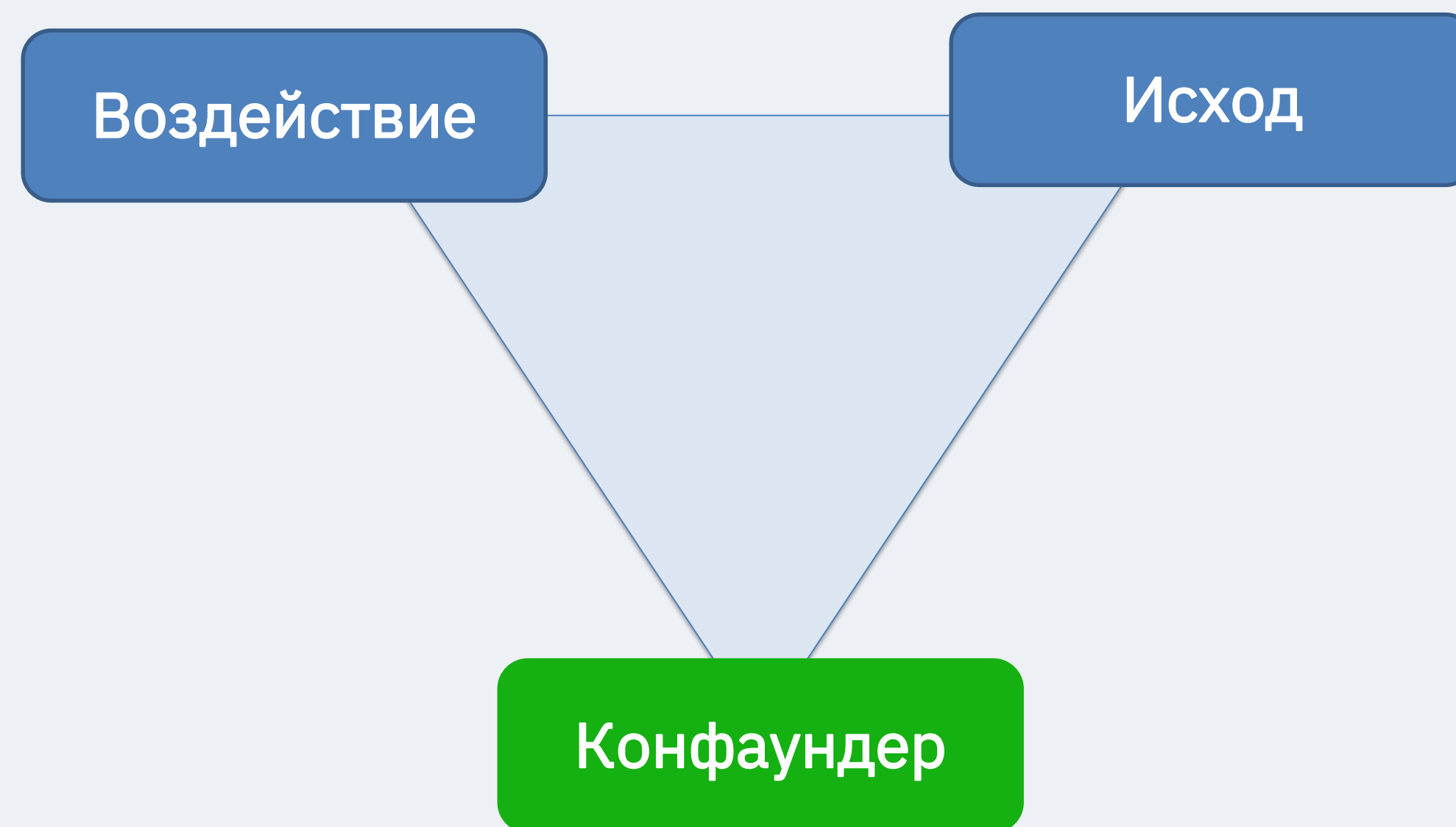
План

1. Конфаундинг-эффект
2. Проблема конфаундинга в исследованиях
3. Подходы к определению потенциальных конфаундеров
4. Методы контроля конфаундинга в исследованиях
5. Использование различных статистических пакетов для контроля конфаундинга (R, StatTech и др.)



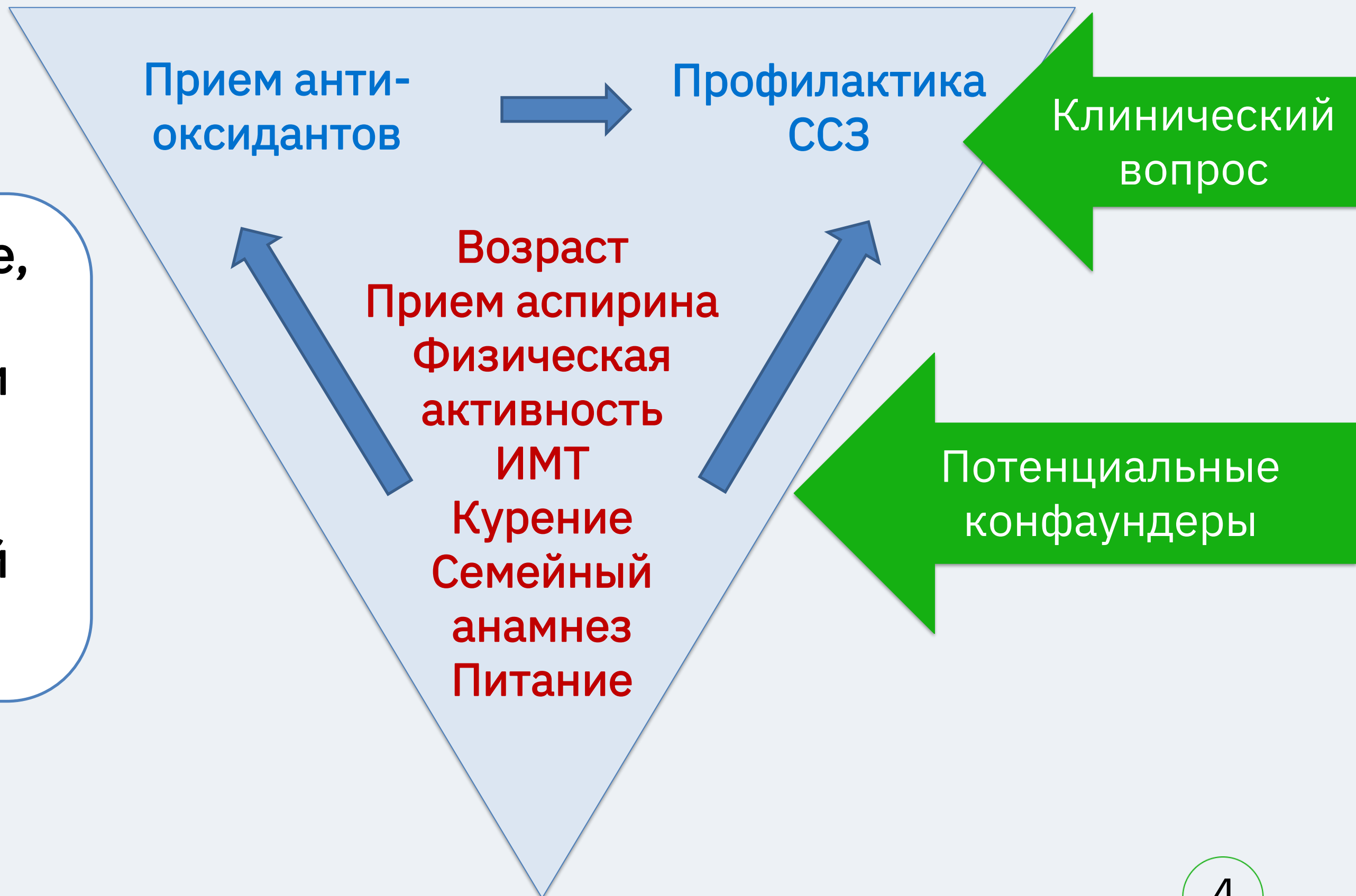
Confounding factor

Вмешивающийся фактор (confounding factor) – фактор, связанный с известным исследуемым фактором (воздействие) и некоторым образом влияющий на результат (исход).



Конфаундинг-эффект

Искажение связи (усиление, ослабление, извращение) между двумя переменными (воздействием и исходом), вызванное присутствием внешнего фактора (третьей переменной).



Проблема конфаундинга в исследованиях

Злоупотребление
алкоголем



Рак легкого

Исследование случай-контроль:
влияние злоупотребления
алкоголем на развитие рака легкого

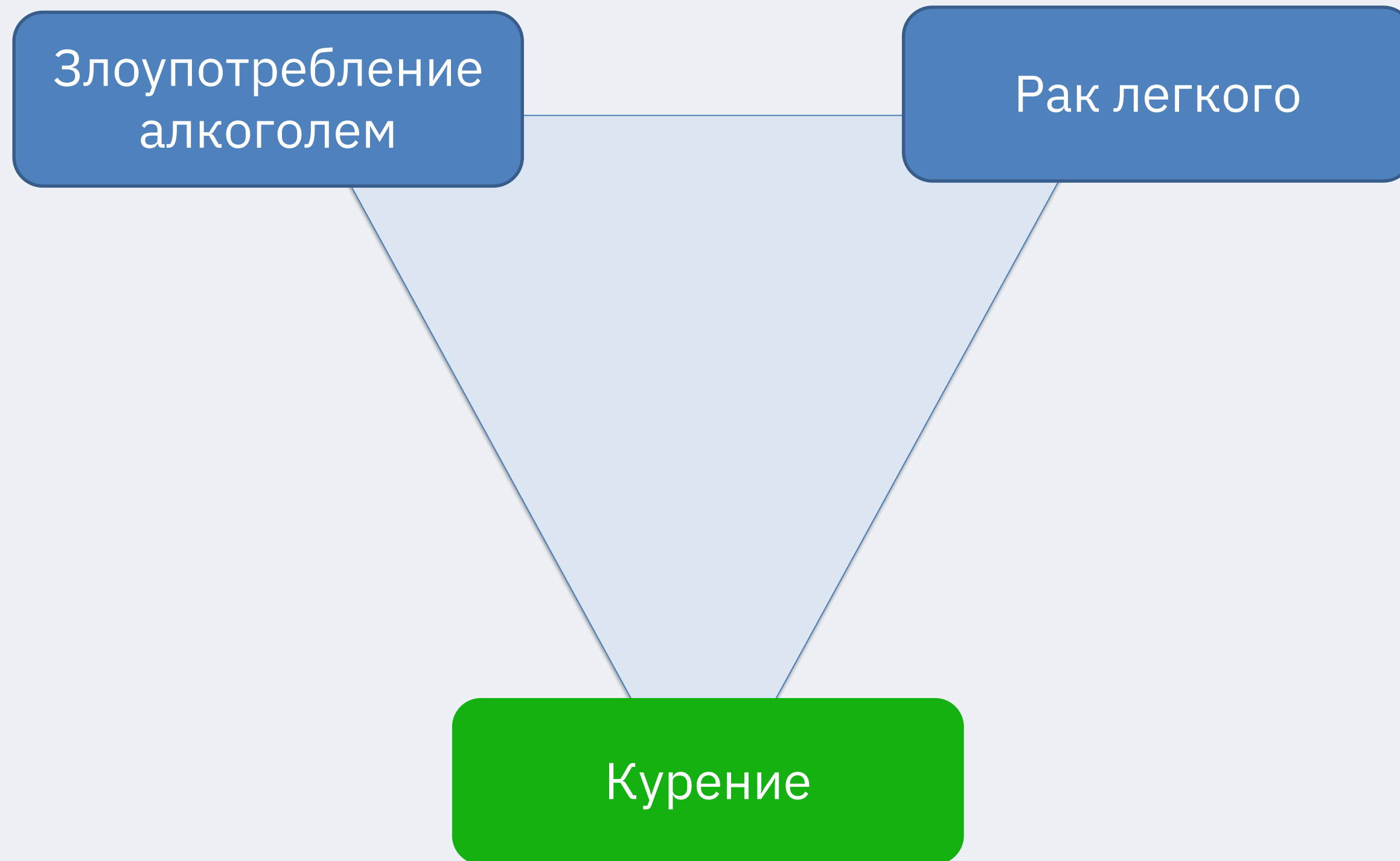
| Cases | Controls |
|-----------|-----------|
| ☹ ☹ ☹ ☹ ☹ | ☺ ☺ ☺ ☺ ☺ |
| ☹ ☹ ☹ ☹ ☹ | ☺ ☺ ☺ ☺ ☺ |
| ☹ ☹ ☹ ☹ ☹ | ☺ ☺ ☺ ☺ ☺ |
| ☹ ☹ ☹ ☹ ☹ | ☺ ☺ ☺ ☺ ☺ |

| | Случаи | Контроли |
|----------|--------|----------|
| Пьющие | 10 | 5 |
| Непьющие | 10 | 15 |

$$OR = (10 * 15) / (10 * 5) = 3$$

Примечание: синие = пьющие, черные = непьющие

Исследование случай-контроль: влияние злоупотребления алкоголем на развитие рака легкого



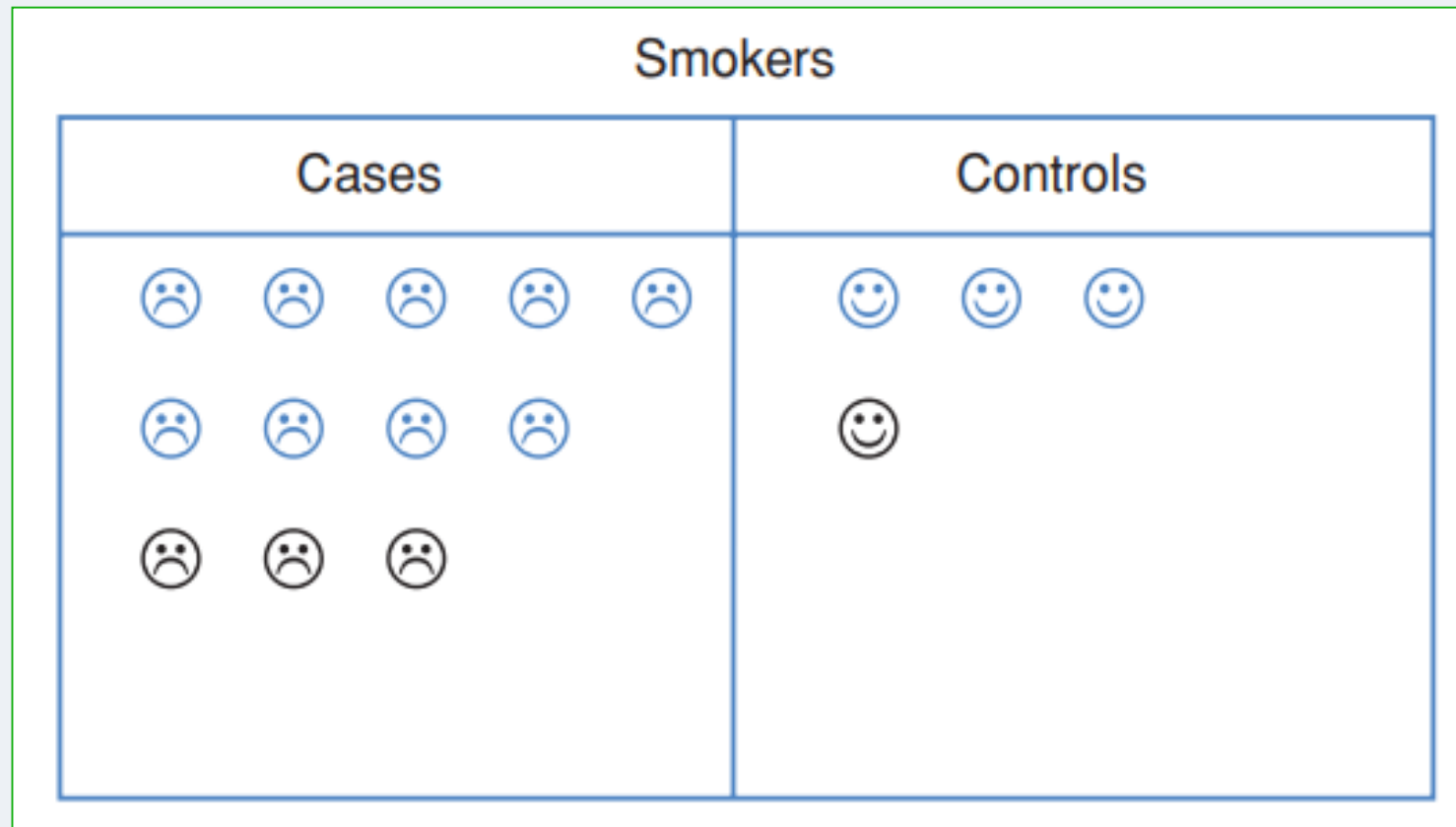
| Smokers | |
|-----------|----------|
| Cases | Controls |
| ☹ ☹ ☹ ☹ ☹ | ☺ ☺ ☺ |
| ☹ ☹ ☹ ☹ | ☺ |
| ☹ ☹ ☹ | |

| Non-smokers | |
|-------------|-----------|
| Cases | Controls |
| ☹ | ☺ ☺ |
| ☹ ☹ ☹ ☹ ☹ | ☺ ☺ ☺ ☺ ☺ |
| ☹ ☹ | ☺ ☺ ☺ ☺ ☺ |
| | ☺ ☺ ☺ ☺ |

синие = пьющие, черные = непьющие

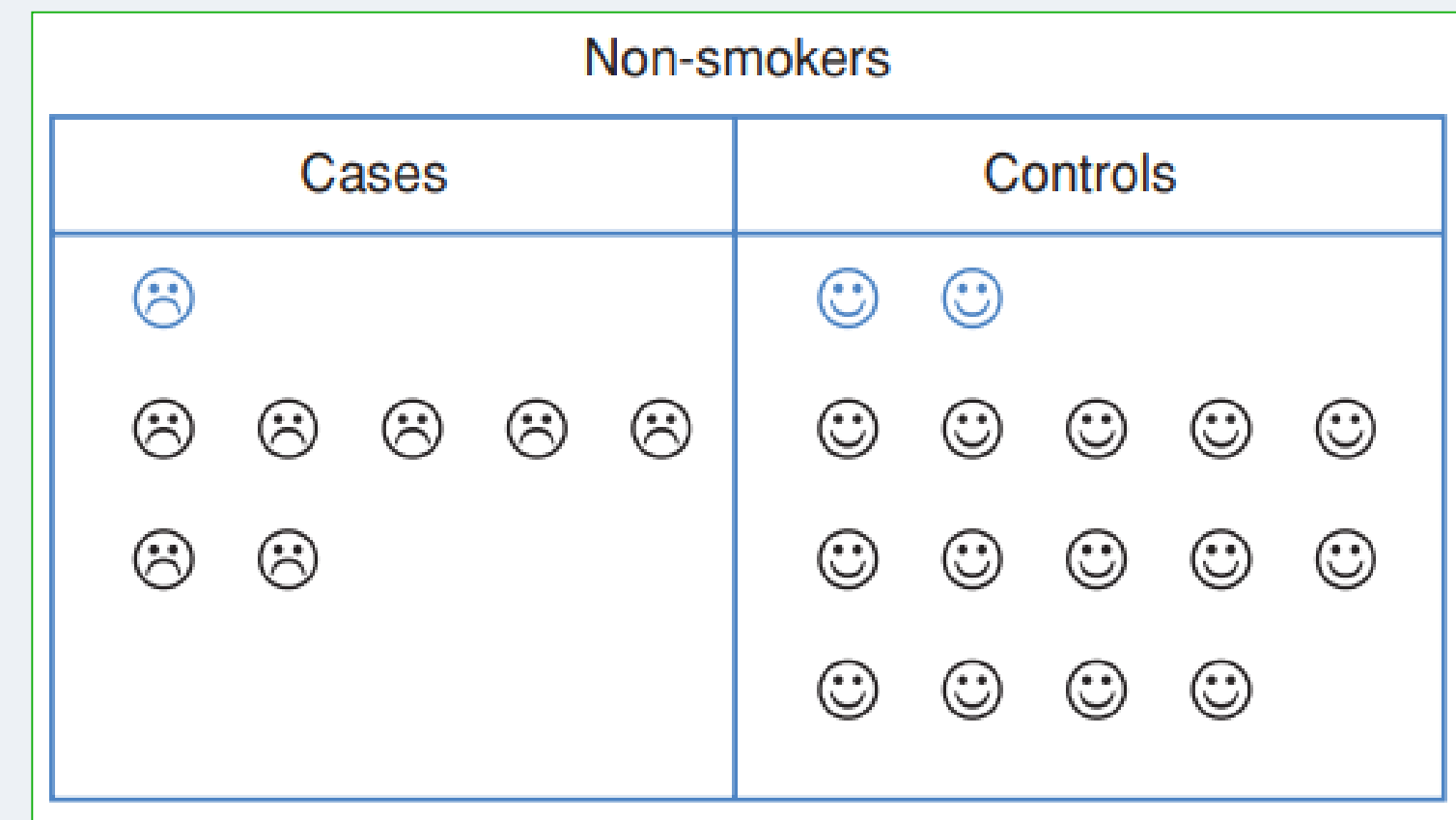


Влияние злоупотребления алкоголем на развитие рака легкого



| | Случаи | Контроли |
|----------|--------|----------|
| Пьющие | 9 | 3 |
| Непьющие | 3 | 1 |

$$OR=(9*1)/(3*3)=1$$

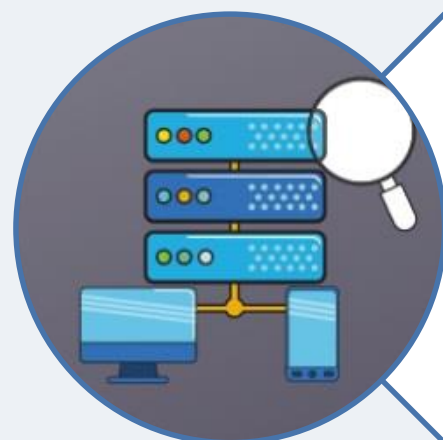


| | Случаи | Контроли |
|----------|--------|----------|
| Пьющие | 1 | 2 |
| Непьющие | 7 | 14 |

$$OR=(1*14)/(2*7)=1$$



Подходы к определению потенциальных конфаундеров



Данные других исследований



Наличие статистической связи с
воздействием и заболеванием



Клинический опыт и знание
биологии заболевания



Методы контроля конфаундинга в исследованиях

Дизайн (планирование)

- Рандомизация
- Ограничение
- Подбор пар

Анализ

- Подбор пар
- Стратификация
- Простая корректировка
- Многофакторная корректировка с использованием регрессии



Использование различных статистических пакетов для контроля конфаундинга (R, StatTech и др.)



Подбор пар (Matching)

| Основная группа | | | | Контрольная | | | |
|--------------------------------------|--------------------|---------|---------|--------------------------------------|-------------------|---------|---------|
| Идентификатор участника исследования | Воздействие | Пол | Курение | Идентификатор участника исследования | Воздействие | Пол | Курение |
| 1a | Высокая активность | Женщина | Нет | 1б | Низкая активность | Женщина | Нет |
| 2a | Высокая активность | Мужчина | Да | 2б | Низкая активность | Мужчина | Да |
| 3a | Высокая активность | Мужчина | Нет | 3б | Низкая активность | Мужчина | Нет |
| 4a | Высокая активность | Женщина | Нет | 4б | Низкая активность | Женщина | Нет |
| И т.д. | | | | И т.д. | | | |



Стратификация и простая корректировка

| Физическая активность | АГ+ | АГ- |
|------------------------------|-----|------|
| Высокая (>8000 шагов в день) | 120 | 1080 |
| Низкая (≤8000 шагов в день) | 324 | 1476 |

Инцидентность в основной (Re) = 10%
 Инцидентность в контрольной (Rne) = 18%
Относительный риск RR = 10%/18%=0,56

| Физическая активность у женщин n=1260 | АГ+ | АГ- |
|--|-----|-----|
| Высокая | 59 | 481 |
| Низкая | 50 | 670 |

Re= 7%
 Rne= 11%
RR = 7%/11%=0,64

| Физическая активность у мужчин n=1740 | АГ+ | АГ- |
|--|-----|-----|
| Высокая | 265 | 995 |
| Низкая | 67 | 413 |

Re= 14%
 Rne= 21%
RR = 14%/21%=0,67



Стратификация и простая корректировка

Мужчины

- RR=0,67
- Доля = 58%, или 0,58

Женщины

- RR=0,64
- Доля = 42%, или 0,42

$$\begin{aligned} RR_{sum.} &= (RR_w * \text{доля женщин}) + (RR_m * \text{доля мужчин}) = \\ &= (0,64 * 0,42) + (0,67 * 0,58) = \mathbf{0,66} \end{aligned}$$

Скорректированный по полу RR = 0,66

Нескорректированный RR = 0,56



Многофакторная корректировка (Multivariable adjustment) с использованием регрессии

Бинарная
логистическая
регрессия

Регрессия Кокса,
или модель
пропорциональных
рисков



Уравнение бинарной логистической регрессии

$$p = \frac{1}{1 + e^{-z}}$$

$$z = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

где:

p – вероятность наступления исхода (в долях единицы),

z – показатель степени в логистической функции,

x – независимые переменные (факторы),

$b_1 \dots b_n$ – коэффициенты логистической регрессии,

a – оценка константы (пересечение),

e – математическая константа (приблизительно равная 2,72).





Многофакторная корректировка с использованием бинарной логистической регрессии

Console Output:

```
Call:
glm(formula = df$цсхп ~ df$частота.употреб.алкоголя + df$сигареты +
df$аллергия + df$прием.стероидов.длительность + df$артериальное.давление +
df$препараты.от.давления + df$сд + df$ибс + df$язв.бол.желудка +
df$аутоимм.заб. + df$нарушение.сна + df$нпвс + df$плс + df$дальнозоркость +
df$близорукость + df$хирург.операции.глаз + df$спорт.пит +
df$бад + df$физ.нагрузка.на.работе, family = binomial(link = "logit"),
data = df)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -1.9272 | -0.7864 | 0.0000 | 0.6158 | 2.0335 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-----------------------------------|-----------|------------|---------|----------|
| (Intercept) | -1.23170 | 0.79815 | -1.543 | 0.1228 |
| df\$частота.употреб.алкоголя | 0.30740 | 0.35110 | 0.876 | 0.3813 |
| df\$сигареты1 | 1.03927 | 0.60876 | 1.707 | 0.0878 |
| df\$аллергия1 | -1.11789 | 0.81605 | -1.370 | 0.1707 |
| df\$прием.стероидов.длительность2 | 1.85918 | 1.59696 | 1.164 | 0.2443 |
| df\$прием.стероидов.длительность3 | 35.38997 | 3292.00839 | 0.011 | 0.9914 |
| df\$прием.стероидов.длительность4 | 2.62189 | 1.68857 | 1.553 | 0.1205 |
| df\$прием.стероидов.длительность5 | -2.91659 | 2.23564 | -1.305 | 0.1920 |
| df\$прием.стероидов.длительность6 | -16.90272 | 3978.11306 | -0.004 | 0.9966 |
| df\$артериальное.давление1 | 3.36798 | 1.31018 | 2.571 | 0.0102 * |
| df\$препараты.от.давления1 | -2.24636 | 1.32261 | -1.698 | 0.0894 |
| df\$сд1 | 16.46492 | 3168.44210 | 0.005 | 0.9959 |
| df\$ибс1 | 0.64334 | 1.13011 | 0.569 | 0.5692 |
| df\$язв.бол.желудка1 | -0.25340 | 1.25577 | -0.202 | 0.8401 |
| df\$аутоимм.заб.1 | -16.88712 | 2207.45897 | -0.008 | 0.9939 |
| df\$нарушение.сна1 | 0.57896 | 0.63474 | 0.912 | 0.3617 |
| df\$нпвс1 | -0.02963 | 0.74353 | -0.040 | 0.9682 |
| df\$плс1 | -5.44598 | 2.27539 | -2.393 | 0.0167 * |
| df\$дальнозоркость1 | 1.18254 | 0.82972 | 1.425 | 0.1541 |
| df\$близорукость1 | -0.94760 | 0.51358 | -1.845 | 0.0650 |
| df\$хирург.операции.глаз1 | -18.07029 | 1870.60613 | -0.010 | 0.9923 |
| df\$спорт.пит1 | 0.99631 | 1.78750 | 0.557 | 0.5773 |
| df\$бад1 | -0.65230 | 1.02557 | -0.636 | 0.5248 |
| df\$физ.нагрузка.на.работе1 | 0.45409 | 0.64277 | 0.706 | 0.4799 |

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Результаты многофакторного анализа в статистических программах



Таблица 1 – Характеристики связи предикторов модели с вероятностью выявления ЦСХП

| Предикторы | Unadjusted | | Adjusted | |
|---|-------------------------|----------|-------------------------|----------|
| | COR; 95% ДИ | p | AOR; 95% ДИ | p |
| Курение | 3,205; 1,545 – 6,646 | 0,002* | 4,103; 1,878 – 8,962 | < 0,001* |
| Прием стероидов в течение последнего месяца | 6,802; 2,705 – 17,099 | < 0,001* | 9,513; 3,615 – 25,028 | < 0,001* |
| Наличие стресса (в сравнении с лицами с низким с уровнем стресса): -Лица с умеренным уровнем стресса | 3,309; 1,799 – 6,086 | < 0,001* | 2,903; 1,489 – 5,658 | 0,002* |
| -Лица с высоким уровнем стресса | 16,666; 1,929 – 144,027 | 0,011* | 25,467; 2,863 – 226,558 | 0,004* |

* – влияние предиктора статистически значимо ($p < 0,05$)

Настройки колонки "ЦСХП" X

Основные Группы Зависимости Модели

Зависимая переменная
ЦСХП

Отбор предикторов
 Пошаговое исключение Принудительное включение

Критерий отбора
 P-value меньше
 Минимальный критерий Акаике

Независимые колонки

- Возраст, лет
- Курение
- Аллергия
- Прием стероидов в течение последнего месяца
- Артериальное давление (нет-0/да-1)
- Сахарный диабет
- Наличие стресса

Таблица 1 – Характеристики связи предикторов модели с вероятностью выявления ЦСХП

| Предикторы | Unadjusted | | Adjusted | |
|---|--------------------------------|--------------------|--------------------------------|--------------------|
| | COR; 95% ДИ | p | AOR; 95% ДИ | p |
| Курение | 3,205; 1,545 – 6,646 | 0,002* | 4,065; 1,837 – 8,989 | 0,001* |
| Аллергия | 0,841; 0,431 – 1,639 | 0,610 | 0,515; 0,221 – 1,198 | 0,123 |
| Прием стероидов в течение последнего месяца | 6,802; 2,705 – 17,099 | < 0,001* | 10,819; 3,975 – 29,459 | < 0,001* |
| Артериальная гипертензия | 1,477; 0,853 – 2,560 | 0,164 | 1,357; 0,685 – 2,689 | 0,382 |
| Сахарный диабет | 5,190; 0,596 – 45,196 | 0,136 | 4,058; 0,380 – 43,337 | 0,246 |
| Наличие стресса (в сравнении с лицами с низким в уровнем стресса): -Лица с умеренным уровнем стресса | 3,309; 1,799 – 6,086 | < 0,001* | 2,816; 1,412 – 5,618 | 0,003* |
| -Лица с высоким уровнем стресса | 16,666; 1,929 – 144,027 | 0,011* | 28,634; 3,152 – 260,343 | 0,003* |

* – влияние предиктора статистически значимо ($p < 0,05$)



Список литературы

1. Эпидемиология: учебник/ В.В. Власов. – 3-е изд., перераб. и доп. – М.: ГЭОТАР-Медиа, 2021. – 496 с.: ил.
2. Наглядная медицинская статистика: учеб. пособие/ Петри А., Сэбин К.; перевод с англ. под ред. В.П. Леонова. – 3-е изд., перераб. и доп. – М.: ГЭОТАР-Медиа, 2015. – 216 с.: ил.
3. Essential epidemiology : an introduction for students and health professionals / Penny Webb and Chris Bain. – 2nd ed., 2011
4. Clinical epidemiology : the essentials / Robert H. Fletcher, Suzanne W. Fletcher, Grant S. Fletcher. – 5th ed., 2014
5. Designing clinical research / Stephen B Hulley . . . [et al.]. – 4th ed., 2013