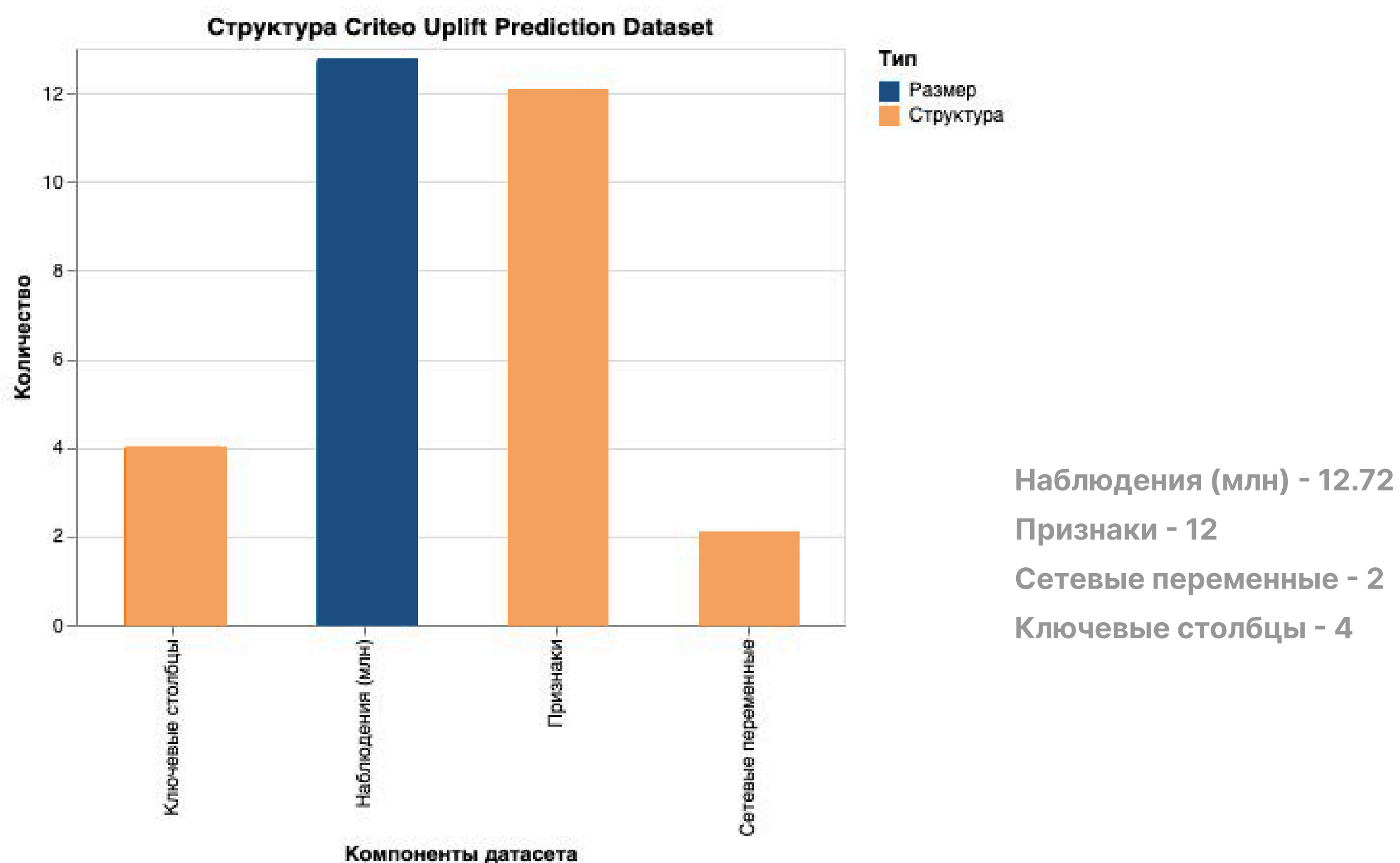


Анализ NWOE и NIV для ATE

Criteo Uplift Prediction Dataset

1. БАЗОВАЯ ИНФОРМАЦИЯ О ДАТАСЕТЕ



Положительные аспекты:

- Большой размер выборки (12.7М наблюдений) обеспечивает высокую статистическую мощность
- Отличный баланс ковариат между группами (все SMD < 0.1, что значительно ниже критического порога 0.25)
- Отсутствие мультиколлинеарности между признаками

Проблемные области:

- Критический дисбаланс групп: 83.9% в тестовой группе vs 16.1% в контрольной (коэффициент 5.21)
- Слабые инструментальные переменные: корреляции visit и exposure с treatment составляют всего 0.025 и 0.082 соответственно

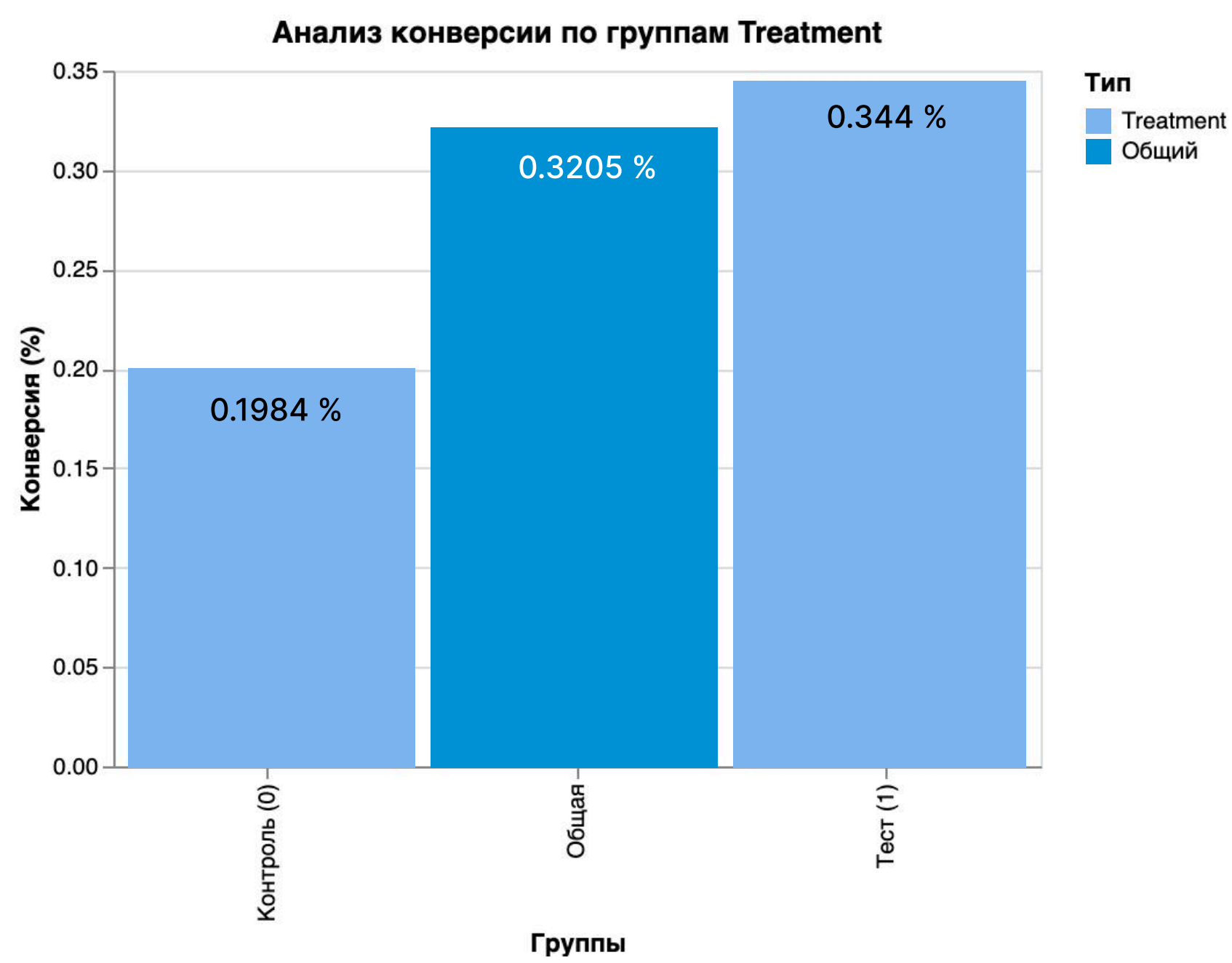
2. АНАЛИЗ TREATMENT ГРУППЫ



Экстремальный дисбаланс (83.9% vs 16.1%) может серьезно повлиять на достоверность оценки АТЕ.

- Снижение статистической мощности для детекции эффектов
- Повышенная чувствительность к выбросам в меньшей группе
- Возможная переоценка эффекта treatment

3. АНАЛИЗ OUTCOME (CONVERSION)

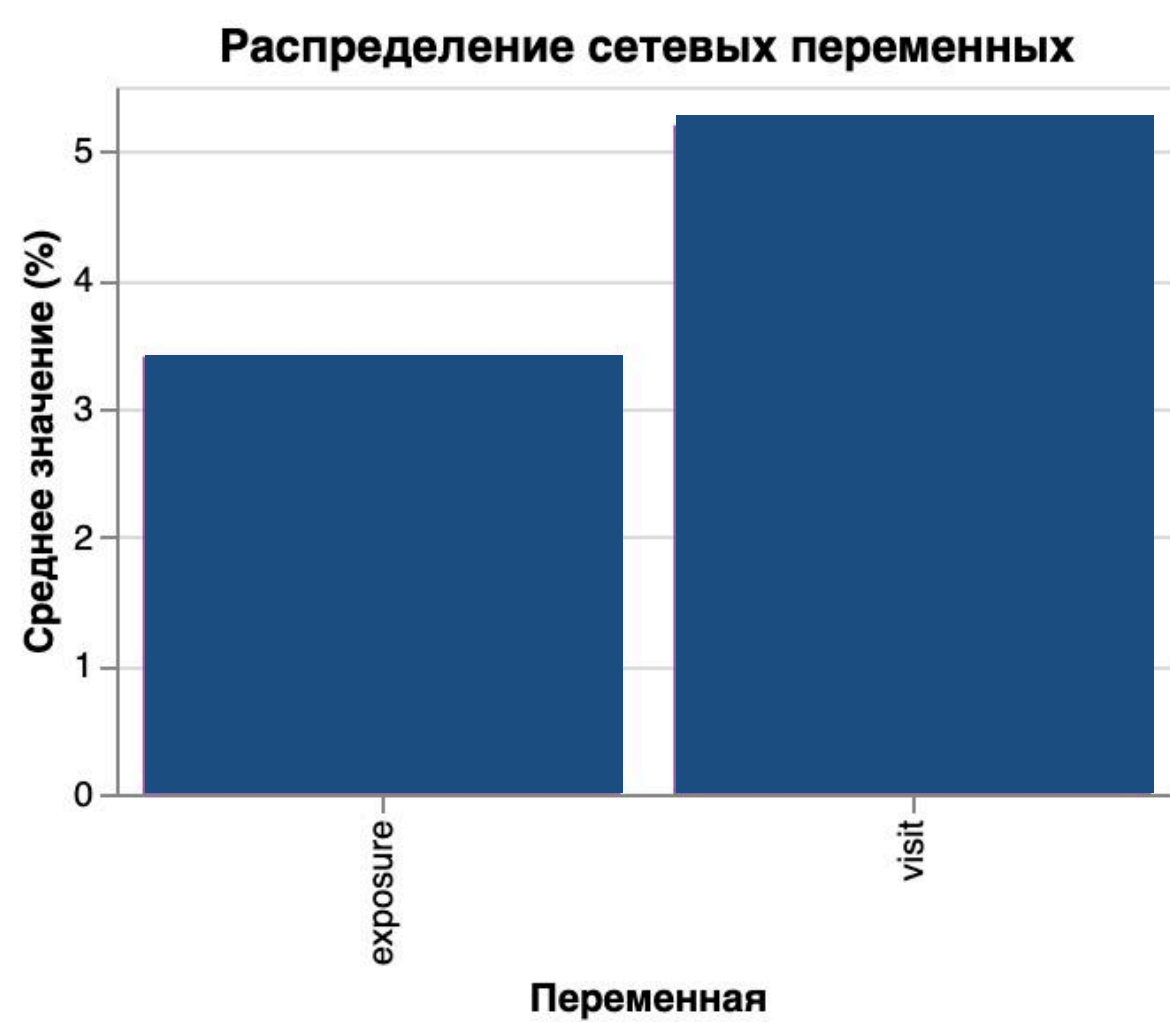


- Очень низкий базовый уровень конверсии (0.32%) - типично для рекламных кампаний
- Treatment группа показывает на 73% выше конверсию (0.34% vs 0.20%)

4. АНАЛИЗ СЕТЕВЫХ ПЕРЕМЕННЫХ

⚠ СЛАБЫЕ ИНСТРУМЕНТАЛЬНЫЕ ПЕРЕМЕННЫЕ:

- visit: корреляция с treatment = 0.025
- exposure: корреляция с treatment = 0.082



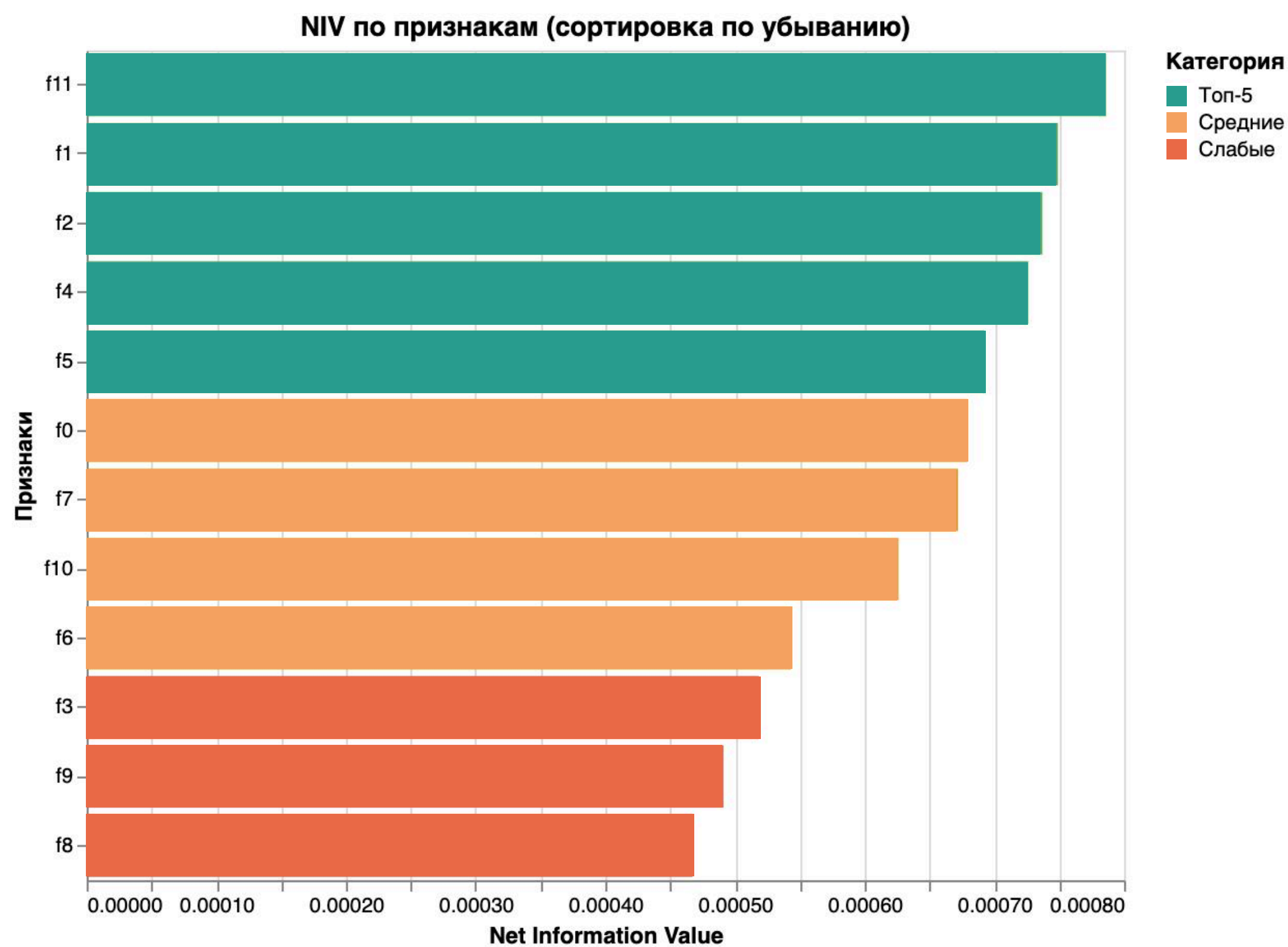
Visit (посещения):

- 5.2% пользователей посетили сайт
- Сильная корреляция с конверсией (0.243) - хороший предиктор
- Слабая связь с treatment (0.025) - плохой инструмент

Exposure (показы рекламы):

- 3.4% получили показы рекламы
- Умеренная корреляция с конверсией (0.167)
- Более сильная связь с treatment (0.082), но все еще слабая

5. NET INFORMATION VALUE (NIV) АНАЛИЗ



Все признаки показывают низкие значения NIV (0.0005-0.0008), что указывает на:

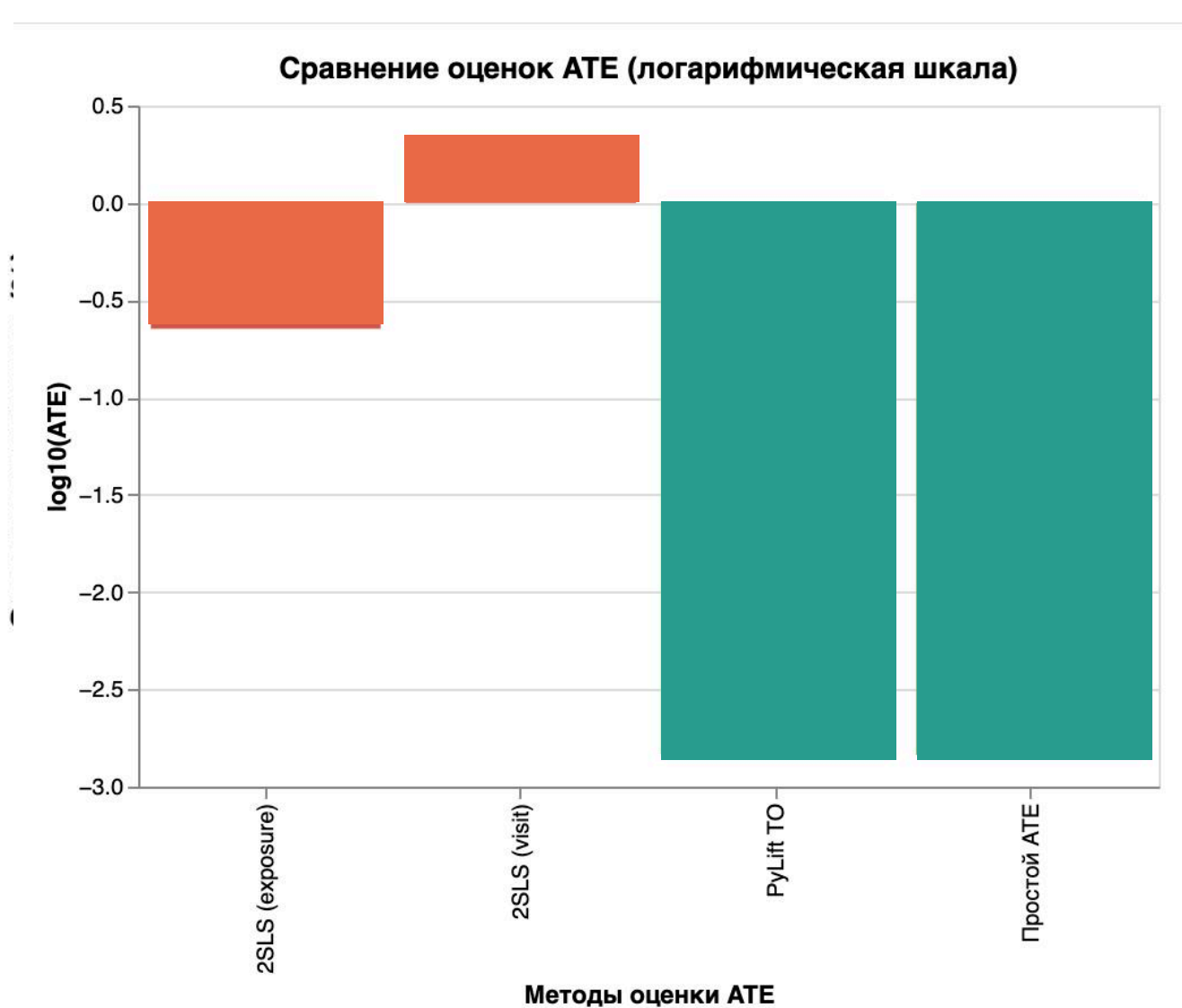
- Слабую предсказательную способность индивидуальных признаков для uplift
- Необходимость использования комбинации признаков для лучшего моделирования

Топ-5 признаков по информативности:

- f11 (0.000786) - наиболее информативен
- f1, f2, f4, f5 - также показывают относительно высокие значения

6. ОЦЕНКИ СРЕДНЕГО ЭФФЕКТА ВОЗДЕЙСТВИЯ (АТЕ)

⚠ 2SLS оценки крайне нестабильны из-за слабых инструментальных переменных. Рекомендуется полагаться на простую оценку АТЕ и Transformed Outcome.



Простой АТЕ = 0.001456
PyLift TO = 0.001470
2SLS (exposure) = 0.225634
2SLS (visit) = 1.959459

1. Простой АТЕ (разность средних): 0.001456

- Базовая оценка без коррекции на ковариаты
- Статистически значим ($p < 0.0001$)
- Может быть смещенной из-за дисбаланса групп

2. PyLift Transformed Outcome: 0.001470

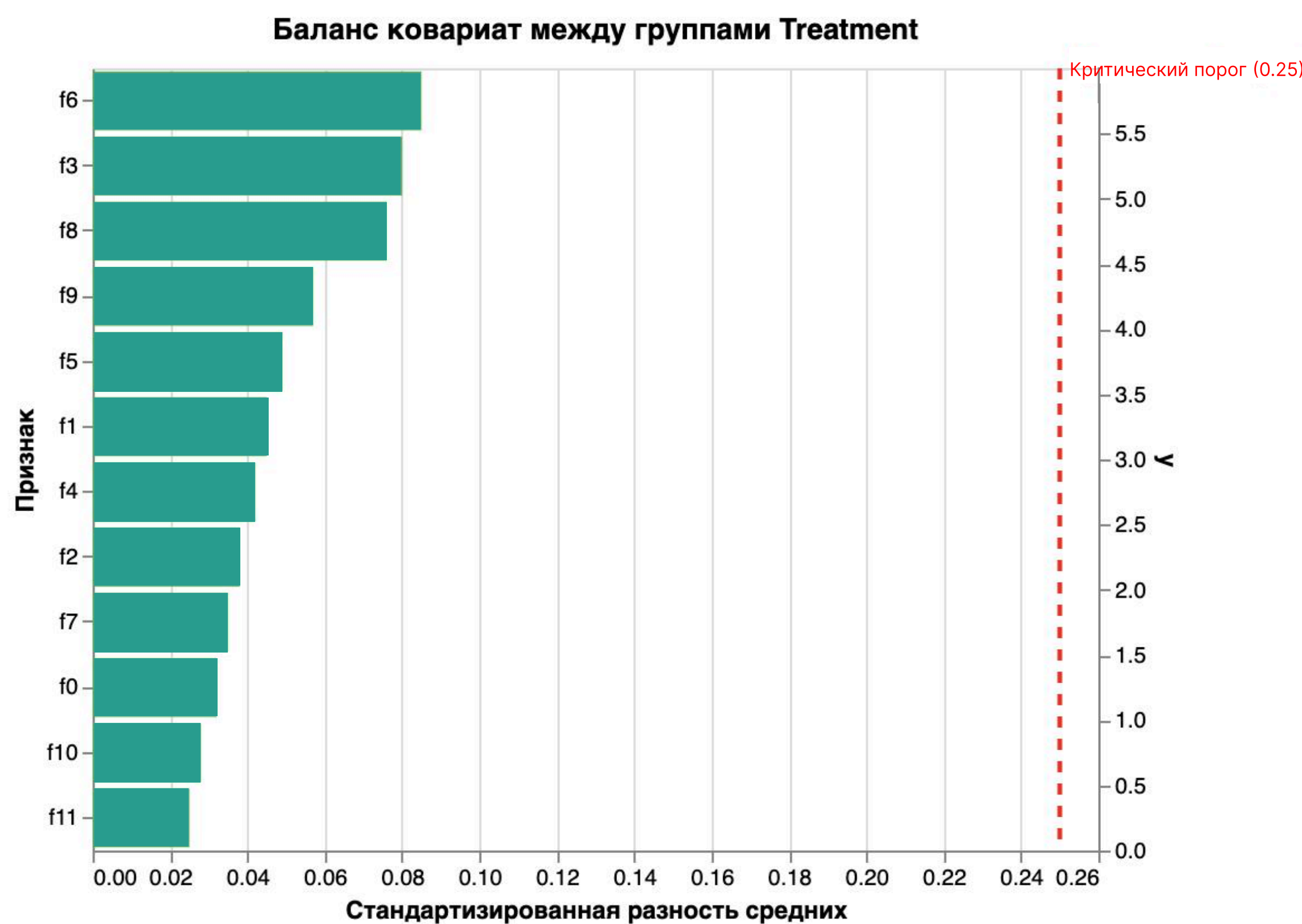
- Очень близко к простой оценке (+1%)
- Учитывает пропенсити скор и ковариаты
- Более надежная оценка

3. 2SLS с инструментальными переменными:

- Visit: 1.959459 - экстремально завышенная оценка
- Exposure: 0.225634 - завышенная в 150+ раз
- Проблема слабых инструментов приводит к нестабильным оценкам

7. БАЛАНС КОВАРИАТ

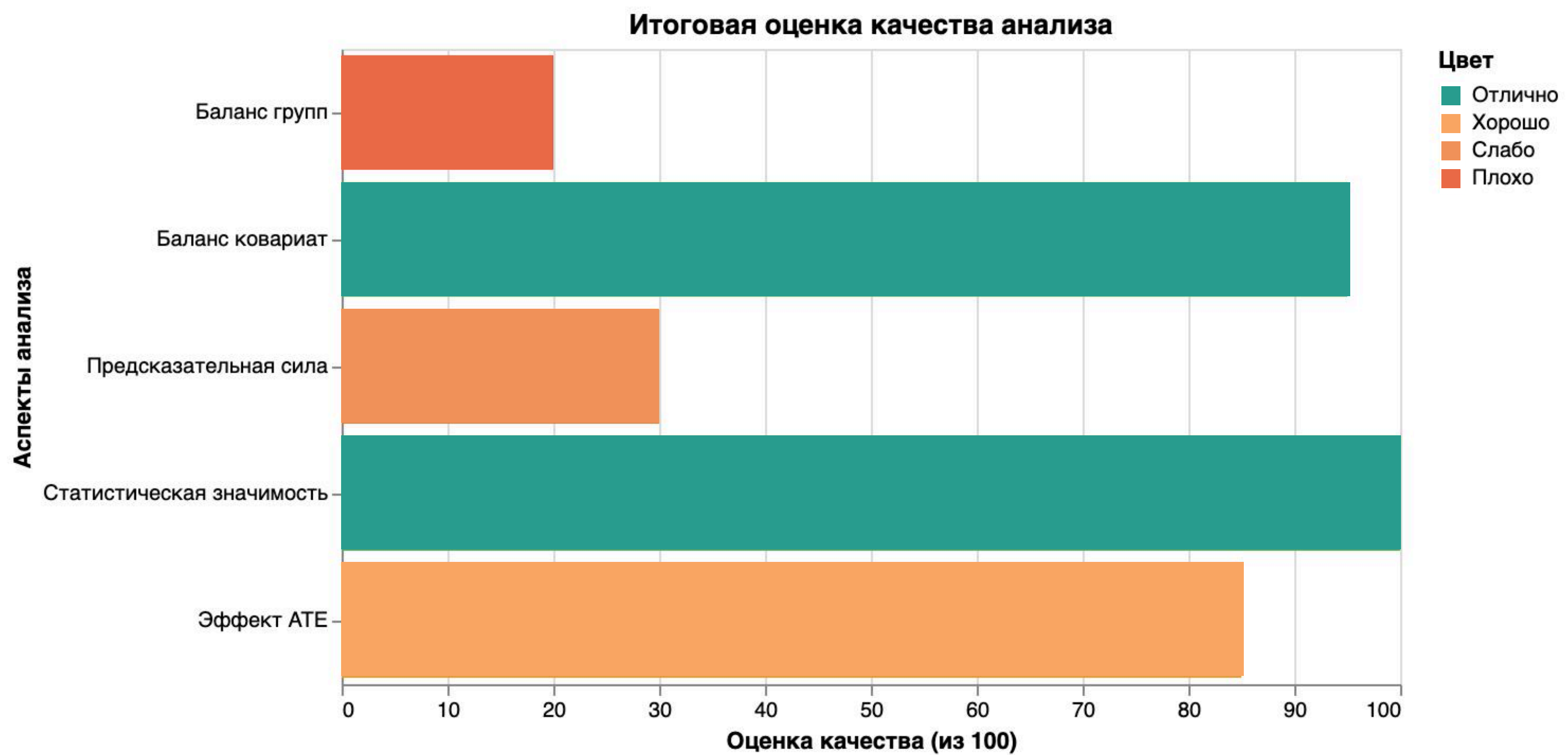
Все признаки имеют $SMD < 0.1$, что указывает на хорошую рандомизацию



Стандартизированная разность средних (SMD) показывает отличный баланс:

- Максимальная $SMD = 0.085$ (f6) - значительно ниже критического порога 0.25
- Большинство признаков имеют $SMD < 0.08$
- Это означает, что группы хорошо сбалансированы по ковариатам
- Дисбаланс treatment не связан с систематическими различиями в признаках

8. Ключевые выводы



Ключевые выводы:

1. Эффект treatment статистически значим: ATE \approx 0.00145-0.00147
2. Относительный прирост конверсии: +73% (с 0.20% до 0.34%)
3. Бизнес-эффект значителен: ~14,600 дополнительных конверсий
4. Признаки f11, f1, f2 наиболее информативны для прогнозирования

- ✓ Использовать оценку ATE \approx 0.00145 как основную
- ✓ Применить методы балансировки для коррекции дисбаланса групп
- ✓ Сосредоточиться на признаках f11, f1, f2 для таргетинга
- ✓ Увеличить размер контрольной группы в будущих экспериментах
- ✓ Поиск более сильных инструментальных переменных для робастности

⚠ Риски

- Дисбаланс групп (5.21:1) может влиять на надежность оценок
- Слабые инструментальные переменные делают 2SLS неприменимым
- Очень низкая базовая конверсия требует больших выборок для точности
- Малые NIV значения указывают на слабую предсказательную силу отдельных признаков