

Ch23. 장기 실험효과 측정

가짜연구소 인과추론팀
온라인 통제 실험 연구자로 거듭나기

2024-06-03

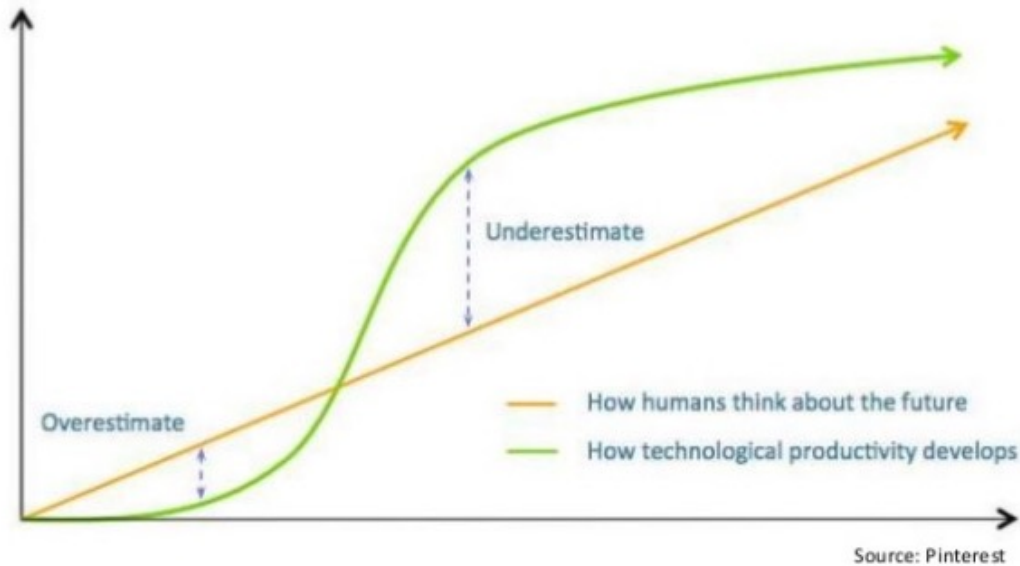
황영산



Causal-Lab

- 실험 중단기 효과가 다른 이유
- 실험 장기 효과를 측정해야 하는 이유
- 구글 광고에서의 장기 효과 측정 및 OEC 개선 사례

What Causes Tech Bubbles? Amara's Law



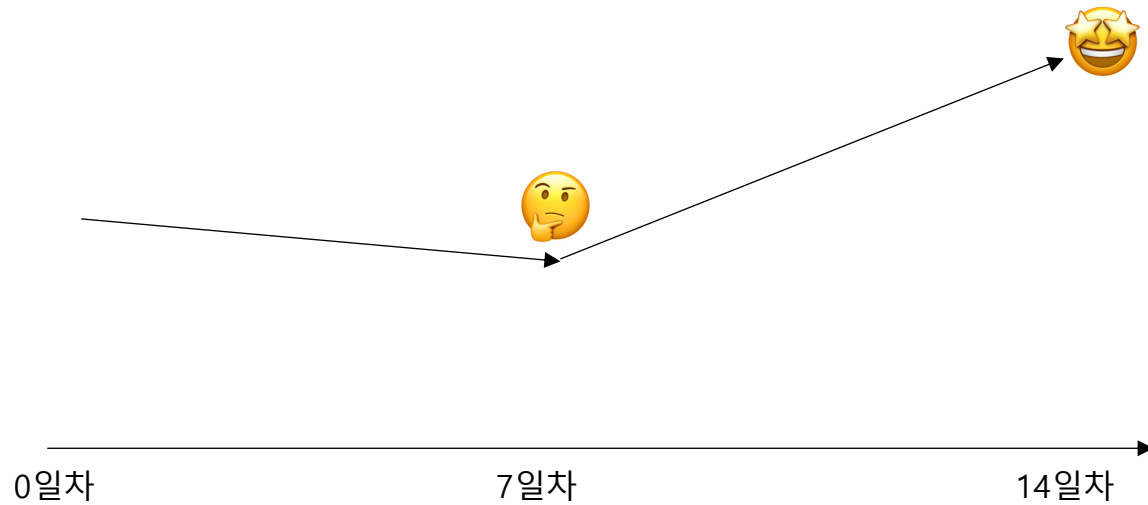
We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run

– Roy Amara

실험 장단기 효과가 다른 이유

1. 사용자 학습 효과(User Learning Effect)

- 사용자가 변화를 학습하고 적응하여 행동이 바뀌는 경우
- 초두 효과(Primacy Effect)와 신기성 효과(Novelty Effect)



초기에 반응이 없는 초두 효과



Performance spikes can sometimes be due to the Novelty Effect

2. 네트워크 효과

- 주변 영향을 받는 커뮤니케이션 앱
- 새로운 기능으로 늘어난 수요와 부족한 공급(ex. 에어비엔비, 이베이, 우버)

3. 지연된 경험 및 측정

- 사용자가 실험 효과를 완전히 경험하기까지 시간차 존재
- 에어비엔비 및 부킹닷컴 같은 경우 온라인 예약 및 실제 목적지 도달까지 간격이 존재
- 연간계약 시, 해당 연도의 누적 경험이 갱신 여부를 결정

4. 생태계 변화

- 다른 새로운 기능 출시
- 계절성
- 경쟁 환경
- 정부 정책
- 컨셉 드리프트
- 오래된 소프트웨어

- 실험의 단기 및 장기 효과의 차이는 ‘**유저 행동**’에서 비롯된다
- 어떠한 유저 행동을 측정할지는 도메인마다 상이하다

- 기업의 장기적인 비전과 전략
 - 기업가치 극대화를 위해서는 장기적인 관점에서 사업, 재무에 관해 전략을 세워야 함
 - 장기 효과를 OEC에 반영함으로써 장기 전략에 기여
- 조직적인 학습
 - 사용자 학습효과(초두 효과, 신기성 효과)로 얻을 수 있는 인사이트
 - 이를 추후 실험에 반영해 향상된 실험 설계
- 일반화
 - 특정 도메인에 대한 원칙을 도출해 이를 일반화하기 위함
 - Google은 검색 광고에서 도출한 원칙을 타사에 적용하였으며 유의미한 결과를 얻음 (*Hohnhold et. al. (2015)*)

Focusing on the Long-term : It's Good for User and Business

Overview

- 단기 효과 != 장기효과
- 장기 효과를 OEC에 반영할 수 있을까?
- Goal : 장기 효과를 반영한 OEC를 만들자!
- Application : Focus on ads blindness & sightedness

Revenue as an OEC

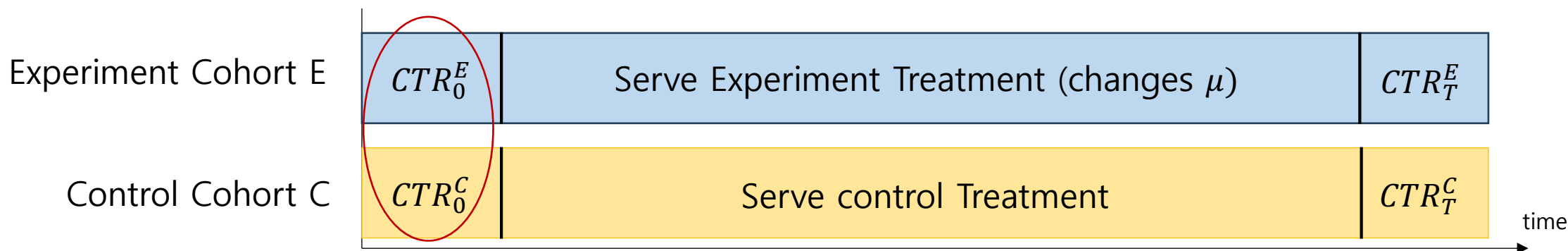
$$\text{Revenue} = \text{Users} * \frac{\text{Queries}}{\text{User}} * \frac{\text{Ads}}{\text{Query}} * \boxed{\frac{\text{Clicks}}{\text{Ad}}} * \frac{\text{Cost}}{\text{Click}}$$

- 장기 효과 유저 행동 변화가 Revenue에 영향을 미침
- 관찰가능한 실험 지표 CTR(Clicks/Ad)에 유저 학습 효과(장기 효과)를 반영한다면?
- 장기 효과를 반영한 Revenue를 계산할 수 있다

Revenue as an OEC

- $CTR_t = \text{Clicks} / \text{Ads} \sim \mu_t * \lambda_t$
 - μ_t : Ad Relevance + UI
 - Changes over short-term
 - λ_t : “User attitude” towards ads(ads blindness & sightedness)
 - Changes over long-term
 - $\Delta\lambda_t > 0$: ads sightedness, $\Delta\lambda_t < 0$: ads blindness
- 장기 효과를 반영한 CTR로 장기 효과 Revenue 계산이 가능!
- Goal
 - 유저 학습 효과인 $\Delta\lambda_t$ 를 측정하자!

Measure : Naive measure



$$\lambda_0^E = \lambda_0^C$$
$$\Delta CTR_0 = \Delta \mu_0$$

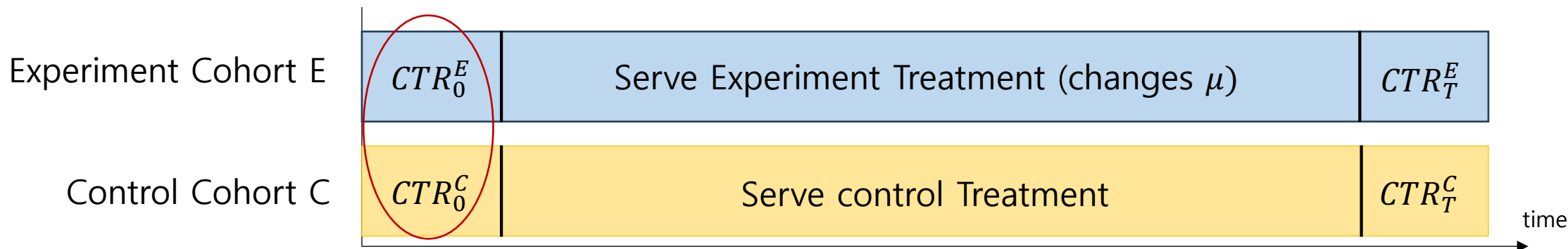
Goal인 $\Delta \lambda_t$ 를 ΔCTR_T 로 구할 수 있을까?

✗ 서비스의 변화가 시간이 지남에 따라 일정하지 않음($\because \Delta \mu_0 \neq \Delta \mu_t$)

Source : [Focusing on the Long-term: It's Good for Users and Business](#)(youtube)

- $CTR_t = \text{Clicks} / \text{Ads} \sim \mu_t * \lambda_t$
 - μ_t : Ad Relevance + UI
 - Changes over **short-term**
 - λ_t : "User attitude" towards ads(ads blindness & sightedness)
 - Changes over **long-term**
 - $\Delta \lambda_t > 0$: ads sightedness, $\Delta \lambda_t < 0$: ads blindness

Measure : Naive measure



$$\lambda_0^E = \lambda_0^C$$
$$\Delta CTR_0 = \Delta \mu_0$$

Goal인 $\Delta \lambda_t$ 를 ΔCTR_T 로 구할 수 있을까?

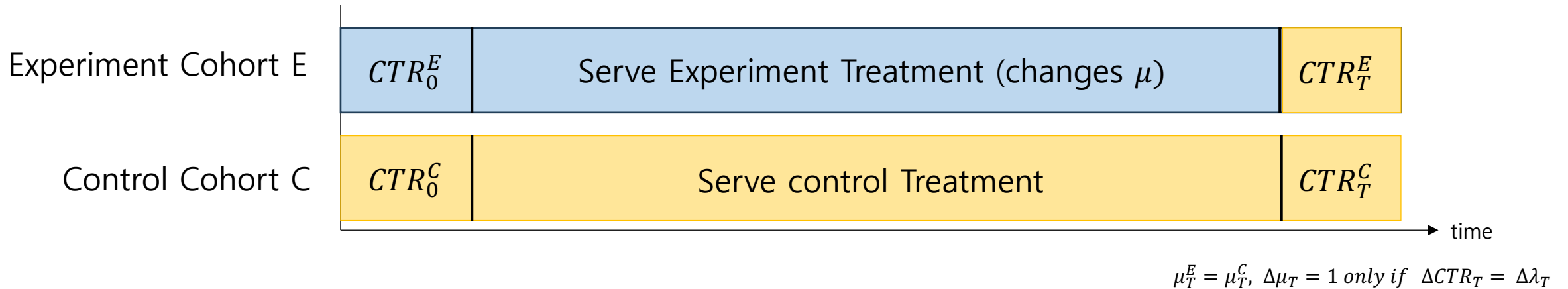
✗ 서비스의 변화가 시간이 지남에 따라 일정하지 않음($\because \Delta \mu_0 \neq \Delta \mu_t$)

그럼 ΔCTR_T 에서 $\Delta \lambda_t$ 를 도출하려면?

$$\Delta CTR_T = \Delta \lambda_T \text{ only if } \Delta \mu_T = 1$$

- $CTR_t = \text{Clicks} / \text{Ads} \sim \mu_t * \lambda_t$
 - μ_t : Ad Relevance + UI
 - Changes over **short-term**
 - λ_t : "User attitude" towards ads(ads blindness & sightedness)
 - Changes over **long-term**
 - $\Delta \lambda_t > 0$: ads sightedness, $\Delta \lambda_t < 0$: ads blindness

Measure : Post-Period Measurement



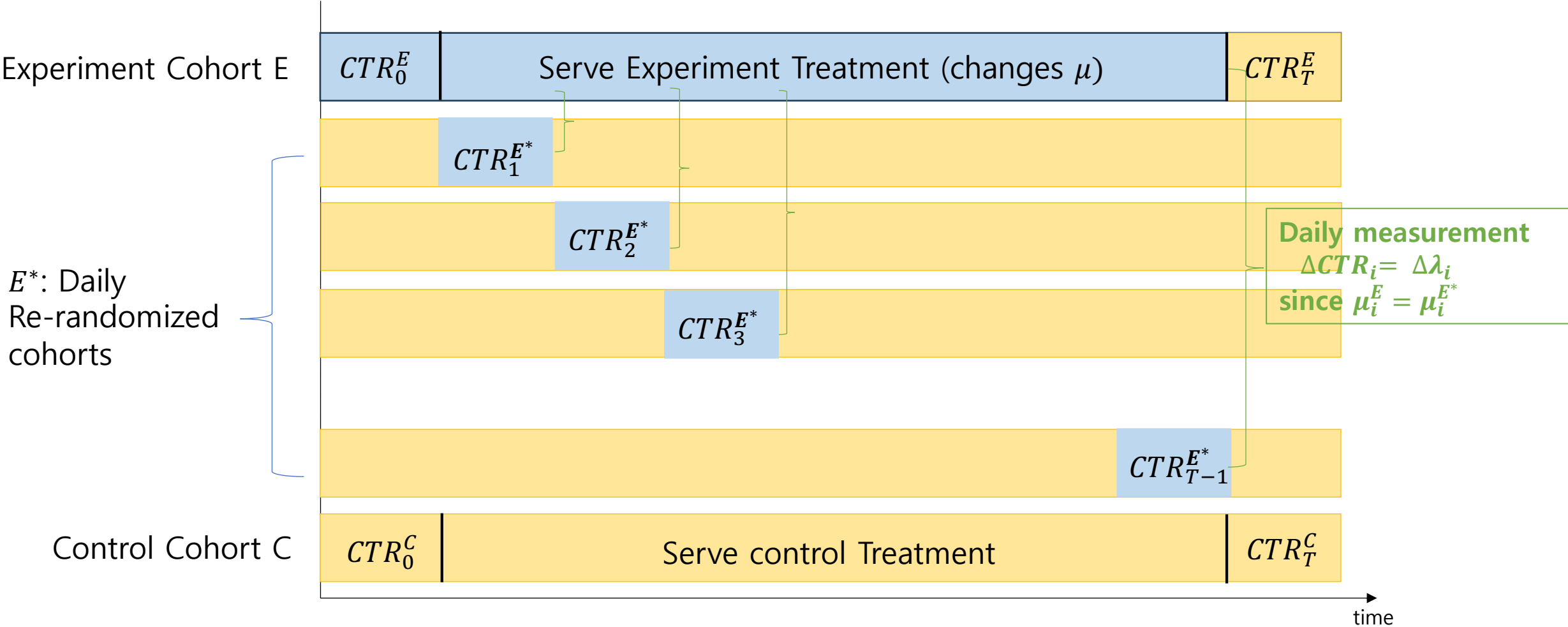
Challenges

- Statistical power
- Takes a long time

Source : [Focusing on the Long-term: It's Good for Users and Business](#)(youtube)

- $CTR_t = \text{Clicks} / \text{Ads} \sim \mu_t * \lambda_t$
 - μ_t : Ad Relevance + UI
 - Changes over **short-term**
 - λ_t : "User attitude" towards ads(ads blindness & sightedness)
 - Changes over **long-term**
 - $\Delta\lambda_t > 0$: ads sightedness, $\Delta\lambda_t < 0$: ads blindness

Measure : Cookie-Cookie-Day



Source : [Focusing on the Long-term: It's Good for Users and Business](#)(youtube)

Modeling : Ads Blindness / Sightedness

$$\Delta\lambda_{\infty} = a * \Delta AdRelevance_0 + b * \Delta LandingPageQuality_0$$

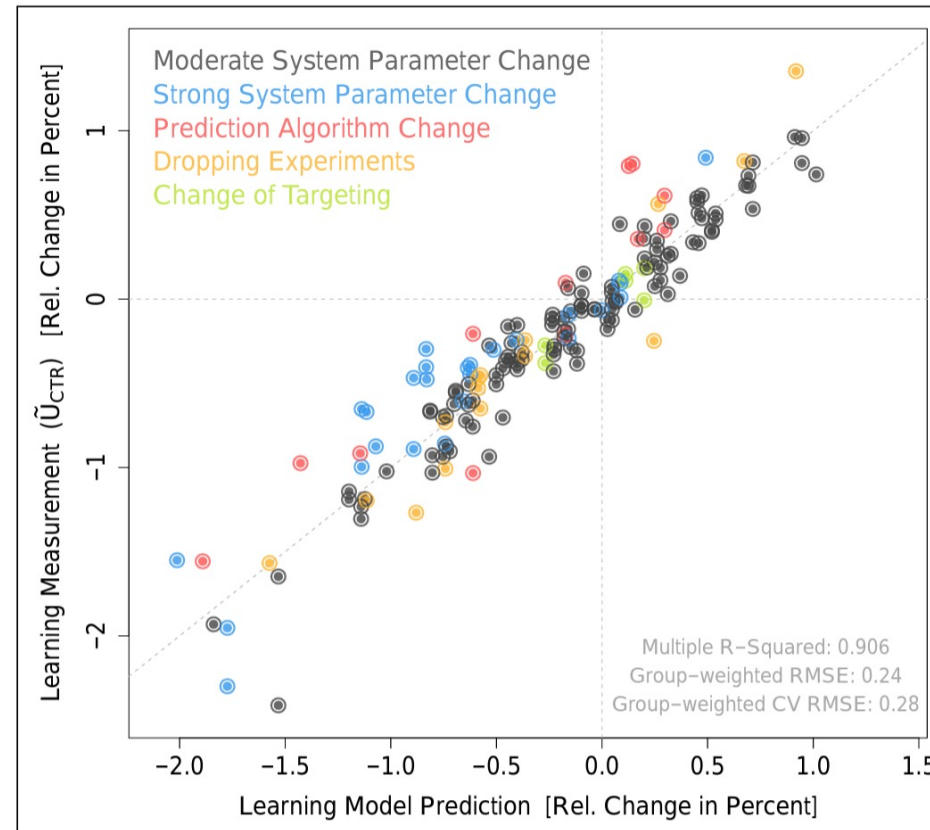


Figure 4: Measured vs. predicted learning for the current desktop macro-model.

Long-term Revenue as an OEC

$$\text{Revenue} = \text{Users} * \frac{\text{Queries}}{\text{User}} * \frac{\text{Ads}}{\text{Query}} * \boxed{\frac{\text{Clicks}}{\text{Ad}}} * \frac{\text{Cost}}{\text{Click}}$$

- 장기 효과 유저 행동 변화가 Revenue에 영향을 미침
- 관찰가능한 실험 지표 $\text{CTR}(\text{Clicks}/\text{Ad})$ 에 유저 행동(장기 효과)를 반영한다면?
- 장기 효과를 반영한 Revenue를 계산할 수 있다

$$\begin{aligned}\Delta LT Rev &\cong \Delta Rev_0 + \Delta \lambda_\infty \\ &\cong \Delta Rev_0 + a * \Delta AdRelevance_0 + b * \Delta LandingPageQuality_0\end{aligned}$$

More ads, more money?

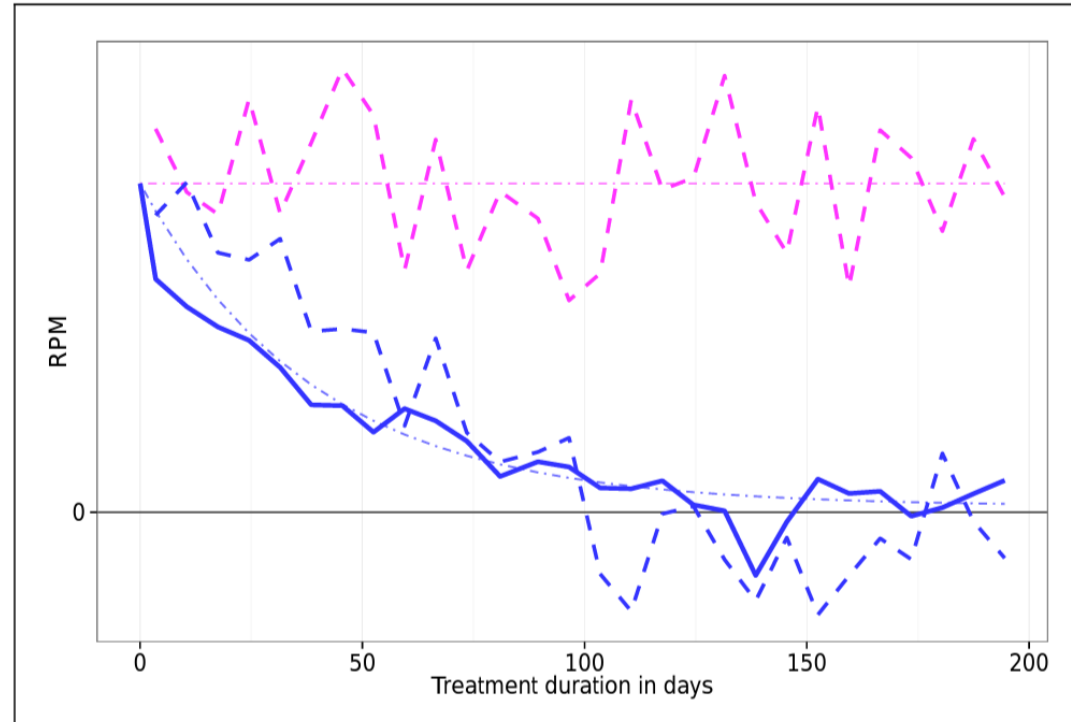


Figure 5: Short-term (pink) and long-term (blue) Δ RPM metrics for simple ad load changes on mobile Google search, restricted to old cookies, 6/26/2013 – 1/9/2014.

- 광고가 많이 표시될수록 단기적 Revenue 증가
- But, 더 많은 광고는 품질 하락으로 이어져 장기적 Revenue는 감소

- 단기 효과 != 장기 효과
- 실험 목적에 따라 OEC에 장기 효과를 반영할 수 있어야한다
- 충분한 장기 실험 결과를 보유하고 있다면 단기 효과로 장기 효과를 예측 및 일반화 할 수 있다
- Focusing on the Long-term : It's Good for User and Business

- 실험 결과에서 초두 효과, 신기성 효과가 보이거나 조치한 사례가 있나요?
- 장기적으로 실험을 설계하거나 결과를 분석 해보셨나요?
- 장기 효과 측정, 유저 학습 효과 계산의 필요성을 느낀적 있나요?