

Given the following causal diagram: **Link at the top -> High CTR**

Treated group: links that happened to be at the top of an article.

Control group: links that happened to not be at the top of an article.

Outcome variable: CTR

We could think of two confounders:

1. Source article length:

Links do not have identical probabilities to be at the top of an article. In fact, its probability depends on the length of the article in which it exists. If the source article is long, the probability of a link to be at the top decreases. At the same time, its probability to have a high CTR decreases also since it would be harder to find it.

In contrast, if the source article is short, the probability of a link to be at the top increases. But also, its probability to have a high CTR increases.

2. Semantic textual similarity between a source article and its target links

if the source article is about a specific topic, the links which are at the top of the page are more likely used to define or explain the topic. Therefore, the semantic similarity between the source article and the links acts as a confounding factor in our study.