Countering Simpson's Paradox with Counterfactuals

Arran Zeyu Wang*, *David Borland*, and *David Gotz
University of North Carolina at Chapel Hill

Introduction

- **Aggregation** is a powerful tool to show summary statistics in visualizations.
- However, it can also introduce additional **risks**, such as **Simpson's Paradox —** *trends that appear at one level of aggregation may disappear or reverse when data is subdivided into lower levels of aggregation.*

Simpson's Paradox

Stone Size	T_A	T_B
Small	**93%** (81/87)	87% (234/270)
Large	**73%** (192/263)	69% (55/80)
All	78% (273/350)	**83%** (289/350)

The above Kidney Stone study included patients with stones of variable size, classified as large or small.

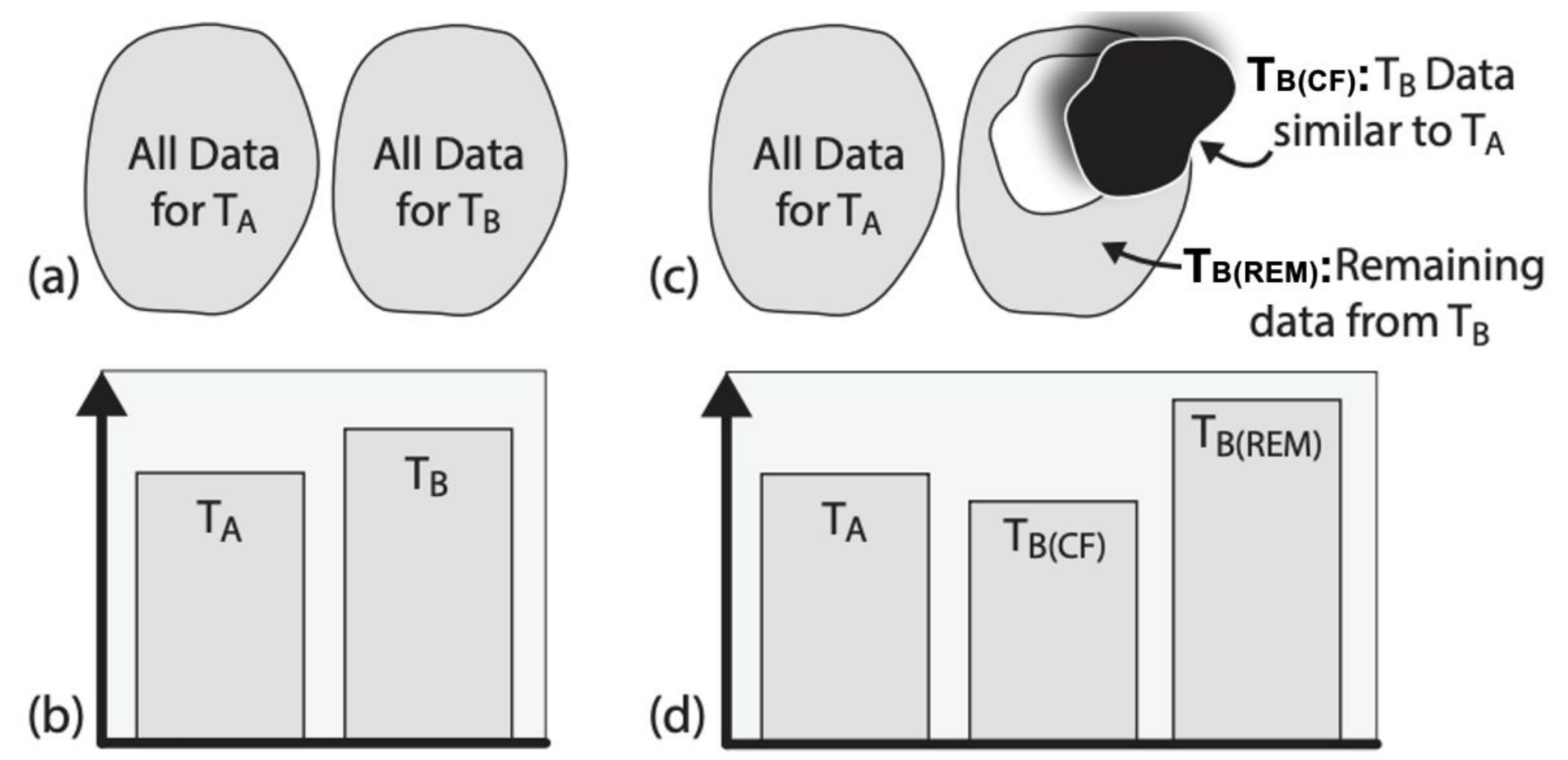
Compared to **Treatment B (*TB*)**, **Treatment A (*TA*)** performed best on small stones and best on large stones. However, counter-intuitively, **Treatment B** appeared to have a higher success rate overall, as a result of the ***unequal distribution*** *of patient groups*.

Visualization of Counterfactuals

Counterfactual reasoning, by constructing **hypothetical scenarios** ("*what if things were the same except for this one fact*?"), can be used to balance the distributions.

Counterfactuals **can be simulated by sampling** from the population receiving ***TB*** a subset of patients similar to those patients receiving ***TA,*** refer to ***TB(CF)***.

TB(CF) will comprise a group of patients with similar variable distributions to ***TA***. In the Kidney example, we sample a group of patients from ***TB*** with the same ratio of *large:small* kidney stones as ***TA*** to include in ***TB(CF)***. The reminder samples are noted as ***TB(REM)***.

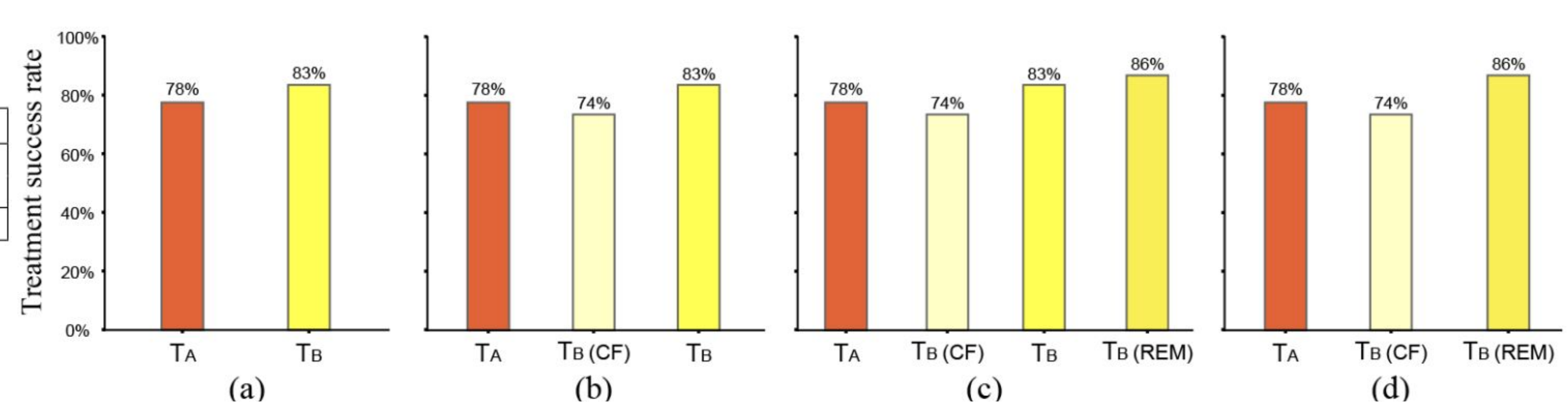


(a-b) are typical visualizations comparing ***TA*** and ***TB*** without considering counterfactuals.
(c) shows the construction of counterfactual subset ***TB(CF)*** and reminder subset ***TB(REM)***.
(d) shows a visualization comparing ***TA*** and ***TB(CF)*** which can avoid Simpson's Paradox.

Countering Simpson's Paradox

Stone Size	T_A	T_B	$T_{B(CF)}$	$T_{B(REM)}$
Small	**93%** (81/87)	87% (234/270)	88% (23/26)	86% (211/244)
Large	**73%** (192/263)	69% (55/80)	69% (55/80)	N/A (0/0)
All	78% (273/350)	**83%** (289/350)	74% (78/106)	86% (211/244)

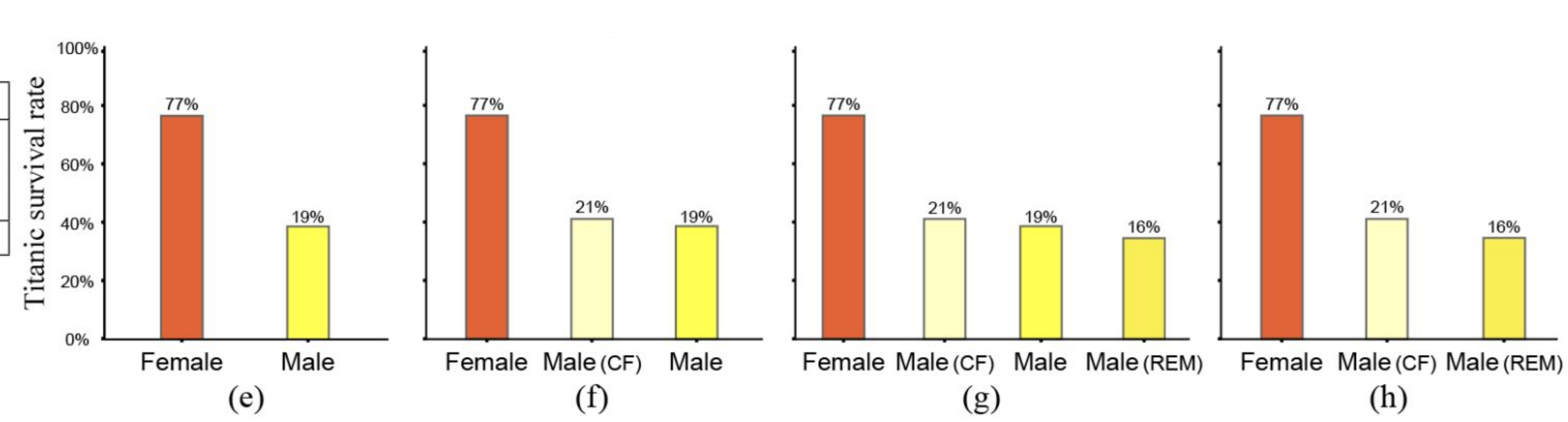
Constructed subsets of Kidney Stone dataset.
w/ Simpson's Paradox



Compared to (a) a traditional visualization of the two treatments, designs that incorporate counterfactuals (b-d) can more accurately communicate the desired comparison (***TB(CF)*** is worse than ***TA***) between treatments.

Cabin	Female	Male	$Male_{(CF)}$	$Male_{(REM)}$
Class 1	**97%** (91/94)	37% (45/122)	37% (35/94)	36% (10/28)
Class 2	**92%** (70/76)	16% (17/108)	16% (12/76)	16% (5/32)
Class 3	**56%** (81/144)	14% (47/347)	14% (20/144)	13% (27/203)
All	**77%** (242/314)	19% (109/577)	21% (67/314)	16% (42/263)

Constructed subsets of Titanic Survival dataset.
w/o Simpson's Paradox



When Simpson's Paradox is not present, a traditional visualization (e) and designs that incorporate counterfactuals (f-h) can both accurately communicate the desired comparison (***Male(CF)*** is lower than ***Female***).

THE VISUAL ANALYSIS and COMMUNICATIONS LABORATORY at UNC
renci
UNC COLLEGE OF ARTS AND SCIENCES Computer Science
UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE