Ensuring anonymity in comparison of institutional performance in large-scale register data using heatmap infused funnel plots

Olsen, Jonas and Wehberg, Sonja

Research Unit for General Practice, Department of Public Health, University of Southern Denmark, Odense, Denmark

2023

Ensuring anonymity in comparison of institutional performance in large-scale register data using heatmap infused funnel plots

Jonas Olsen, Sonja Wehberg[1]

[1]Research Unit for General Practice, Department of Public Health, University of Southern Denmark, Odense, Denmark

Abstract
Comparing institutional performance in large-scale register studies while ensuring anonymity poses a challenge in terms of privacy preservation and accuracy of published results. This paper introduces a novel approach that combines heatmaps and funnel plots to address this challenge. Heatmap-infused funnel plots provide a visual representation of institutional performance while protecting privacy of individuals or institutions.

Keywords:
Funnelcompar, Heatplot, Epidemiology, Methods, Quality Assurance, Health Care Outcome Assessment, Health Care

---

1 Introduction

Comparison of institutional performance plays a crucial role in various domains, including healthcare, education, and finance. Traditional methods of comparing institutions such as league tables often overlook the inherent variability in their performance, leading to biased assessments (1). Funnel plots, where institutions are plotted based on the measure of interest and its precision, have been proposed as a solution (2). However, anonymity of institutional performance can be compromised. This might especially pose a challenge when working in research environment systems such as *Statistics Denmark*; these systems typically provide access to microlevel nationwide register data but require full anonymity for all data extraction.

In this paper we introduce a novel approach that integrates heatmaps into funnel plots. We show how heatmap-infused funnel plots offer a visually intuitive representation of performance variability, while ensuring anonymity.

2 Methods

2.1 Data

We simulated a dataset with 750 institutions, each assigned a random number of individuals from a normal distribution with a mean of 250 and standard deviation of 50. Each individual institution was assigned a uniformly distributed random number between 0-1 as our measure of interest. The simulated data could represent surgery complications among patients in specific hospitals or test scores among children in schools.

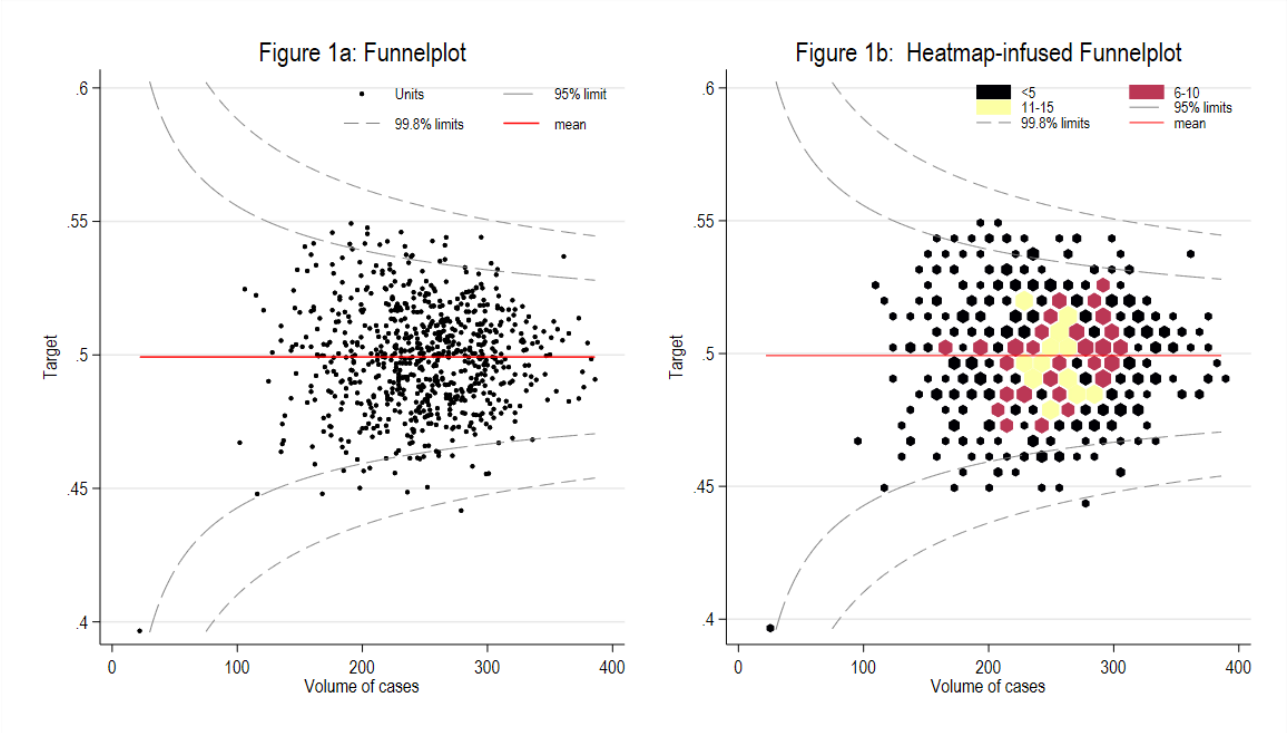
2.2 Statistics

We generated a funnel plot using the user written Stata command *funnelcompar (3)* with control limits set at 2 and 3 standard deviations from the mean, following the approach proposed by Spiegelhalter (2) (figure 1a).

To represent the local concentration of units, we created a heatmap using the user written Stata command *heatplot* (4). The heatmap employed a colour gradient and symbol size to the local concentration of units (figure 1b). All calculations and graphics were performed in StataCorp. 2021. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.

3 Example

Figure 1a displays a funnel plot of our simulated data, providing an overview of the relationship between institutional performance and measurement precision. However, the possible identification of institutions from the graph hinders publication according to article 32 of the General Data Protection Regulation (5). In contrast, the heatmap infused funnel plot in figure 1b preserves the precision of measurement while ensuring anonymity.



4 Conclusion

The integration of heatmaps into funnel plots offers a novel and effective approach for anonymizing large-scale register data while preserving comparability of institutional performance. The proposed methodology successfully tackles the challenges associated with privacy preservation and accuracy. By employing heatmap-infused funnel plots, researchers can gain valuable insights into institutional performance while respecting privacy regulations.

Acknowledgements We would like to thank Benn Jann for his work on the user written Stata command *heatplot* and Silvia Forni and Rosa Gini for their work on the user written Stata command *funnelcompar*.

Availability of data The Stata do-file is provided as supplementary material.

Author contributions All authors contributed to the conception of the method. The first draft of the manuscript and code was written by Jonas Olsen. All authors commented and provided revisions on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval The study is based purely on simulated data. No approval was needed.

Conflicts of interest The authors declares no conflict of interest.

References

1. Goldstein H, Spiegelhalter DJ. League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. Journal of the Royal Statistical Society Series A (Statistics in Society). 1996;159(3):385-443.
2. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Statistics in Medicine. 2005;24(8):1185-202.
3. Forni S, Gini R. FUNNELCOMPAR: Stata module to perform funnel plot for institutional comparison. https://ideas.repec.org/c/boc/bocode/s457078.html: Boston College Department of Economics; 2009.
4. Jann B. HEATPLOT: Stata module to create heat plots and hexagon plots. https://ideas.repec.org/c/boc/bocode/s458598.html: Boston College Department of Economics; 2019.
5. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 02016R0679 — EN — 04.05.2016 — 000.002 (2016).

About the authors:

Jonas Olsen is a medical doctor and phd-student at the Research Unit for General Practice at University of Southern Denmark.

Sonja Wehberg is a statistician at the Research Unit for General Practice at University of Southern Denmark.

Supplementary material Stata code

```
/***********************************
Heatmap-infused funnel plots - supplementary STATA do-file
Author: Jonas Olsen
Date: 20jun2023
Description: STATA code used to produce simulated data and figures
in the article
"Ensuring anonymity in comparison of institutional performance in large-
scale
register data using heatmap infused funnel plots"

Requried packages:
*Funnelcompar by Silvia Forni and Rosa Gini
ssc install funnelcompar
*Heatplot by Ben Jann
ssc install heatplot
```

```
**********************************/

clear
set seed 3636

set obs 750
gen institution=_n
gen npt=trunc(rnormal(250,50))
expand npt
gen value=runiform()
bysort institution: egen mean_value=mean(value)
bysort institution: egen sd_value=sd(value)

bysort institution: keep if _n==1

funnelcompar mean_value npt institution sd_value, conti display nodraw
local target_val=`r(target_val)'
local target_sd=`r(target_sd)'

twoway  ///
          ||  scatter mean_value npt, mc(black) msize(tiny) ///
          ||  function `target_val'+1.959963984540054*`target_sd'/sqrt(x)
, range(30 386) lc(gs8) lp(longdash) lw(thin)  ///
          ||  function `target_val'-1.959963984540054*`target_sd'/sqrt(x)
, range(30 386) lc(gs8) lp(longdash) lw(thin)    ///
          ||  function `target_val'+3.090232306167813*`target_sd'/sqrt(x)
, range(75 386) lc(gs8) lp(dash) lw(thin)    ///
          ||  function `target_val'-3.090232306167813*`target_sd'/sqrt(x)
, range(75 386) lc(gs8) lp(dash) lw(thin)    ///
          ||  function `target_val' ,lcolor(red) range(npt) lstyle(solid)
xscale() ylabel(, angle(   horizontal))  ///
          xtitle(`"Volume of cases"') ytitle(`"Target"') title(`"Figure
1a: Funnelplot"')   ///
          legend(order( 1 `"Units"' 2 "95% limit" 4 "99.8% limits" 6
"mean" ) symxsize(*0.5) cols(2) pos(2) ring(0) size(small)
region(lstyle(none) fcolor(none))) graphregion(c(white))
name(funnel_plot, replace)

graph export ".../funnel.png", as(png) replace

*Heatplot
funnelcompar mean_value npt institution sd_value, conti display nodraw
local target_val=`r(target_val)'
local target_sd=`r(target_sd)'

hexplot mean_value npt, statistic(count) colors(inferno) cut(1(5)@max)
size srange(0.1) keylabels(,range(1))  ///
          legend(order(1 "<5" 2 "6-10" 3 "11-15" 4  "95% limits" 6 "99.8%
limits" 8 "mean") region(lstyle(none) fcolor(none)) ///
          cols(2) pos(2) ring(0) subtitle("") size(small) symxsize(*1.5)
symysize(*0.75) ) addplot( ///
          function `target_val'+1.959963984540054*`target_sd'/sqrt(x) ,
range(30 386) lc(gs8) lp(longdash) lw(thin)  ///
```

```
          ||  function `target_val'-1.959963984540054*`target_sd'/sqrt(x)
, range(30 386) lc(gs8) lp(longdash) lw(thin)    ///
          ||  function `target_val'+3.090232306167813*`target_sd'/sqrt(x)
, range(75 386) lc(gs8) lp(dash) lw(thin)    ///
          ||  function `target_val'-3.090232306167813*`target_sd'/sqrt(x)
, range(75 386) lc(gs8) lp(dash) lw(thin)    ///
          ||  function `target_val' ,lcolor(red) range(npt) lstyle(solid)
lw(thin) ) ///
          name(hexplot, replace) ylabel(, angle(   horizontal))
xtitle(`"Volume of cases"') ytitle(`"Target"') title(`"Figure 1b:
Heatmap-infused Funnelplot"') graphregion(c(white) )

graph export ".../heatmap_funnel.png", as(png) replace

graph combine funnel_plot hexplot, ysize(4) xsize(7) graphregion(c(white)
)
graph export ".../combined_heatmap-infused_funnel.png", as(png) replace

exit
```