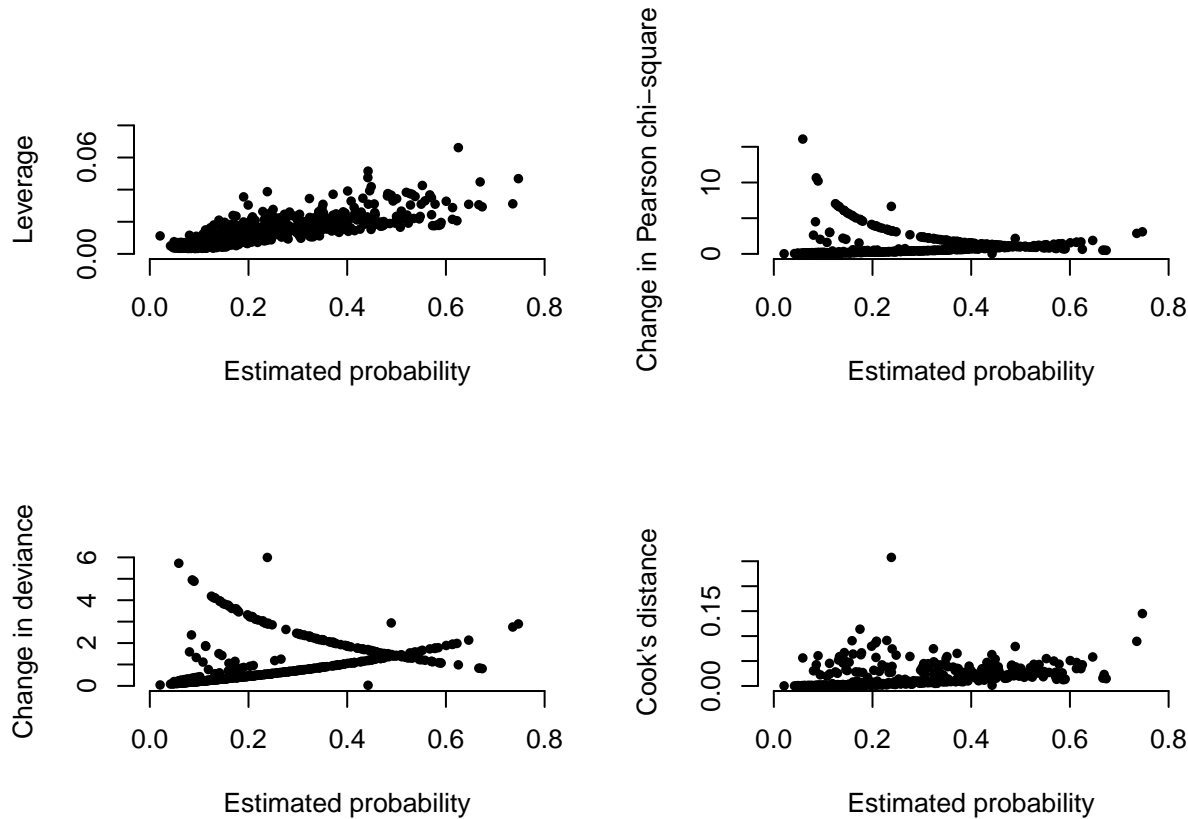


# Inconsistencies in the reproduction of figures 5.12 and 5.15 of Hosmer et al. 2013

Johann Popp

13 October 2017

Hosmer et al<sup>1</sup> have suggested several graphics to identify and investigate extreme and influencing covariate patterns in logistic regression models. These plots were reproduced using the computer software R<sup>2</sup>, taking data from package `aplore3`<sup>3</sup> and using package `epiR`<sup>4</sup> for the calculation of statistics based on covariate patterns rather than single data rows.



This works fine for the graphs showing change of Pearson chi-square and change of deviance but the plots of leverage vs. fitted values and Cook's distance vs. fitted values differ substantially from figures 5.12 and 5.15 of the book.

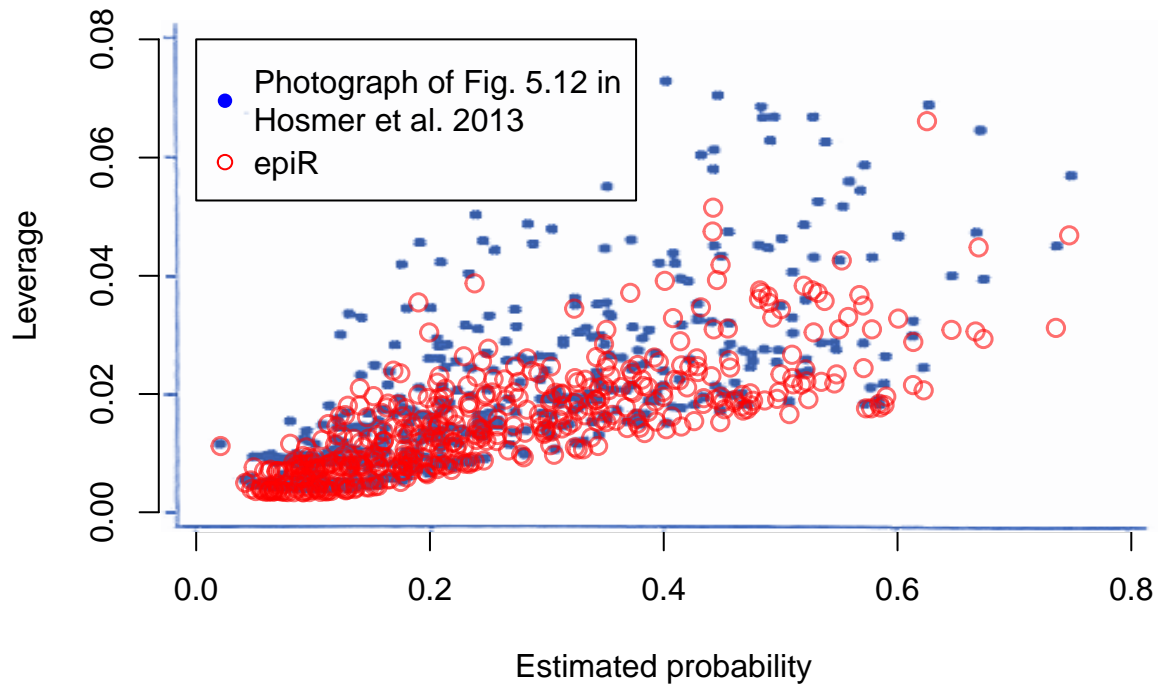
<sup>1</sup>Hosmer, David W., Stanley Lemeshow, und Rodney X. Sturdivant. Applied logistic regression. 3rd Ed. Wiley series in probability and statistics. Hoboken, NJ: Wiley, 2013, p 186 ff.

<sup>2</sup>R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

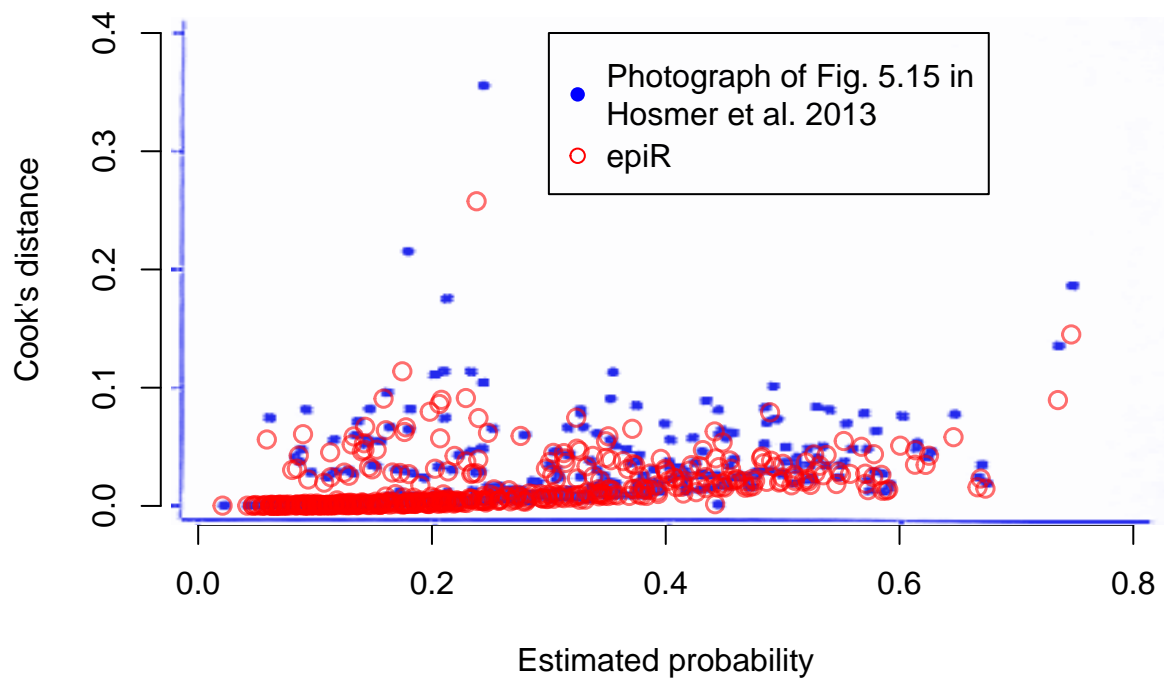
<sup>3</sup>Luca Braglia (2016). `aplore3`: Datasets from Hosmer, Lemeshow and Sturdivant, "Applied Logistic Regression" (3rd Ed., 2013). R package version 0.9. <https://CRAN.R-project.org/package=aplore3>

<sup>4</sup>Mark Stevenson with contributions from Telmo Nunes, Cord Heuer, Jonathon Marshall, Javier Sanchez, Ron Thornton, Jenő Reiczig, Jim Robison-Cox, Paola Sebastiani, Peter Solymos, Kazuki Yoshida, Geoff Jones, Sarah Pirikahu, Simon Firestone and Ryan Kyle. (2017). `epiR`: Tools for the Analysis of Epidemiological Data. R package version 0.9-87. <https://CRAN.R-project.org/package=epiR>

### Leverage based on package epiR vs. published figure 5.12

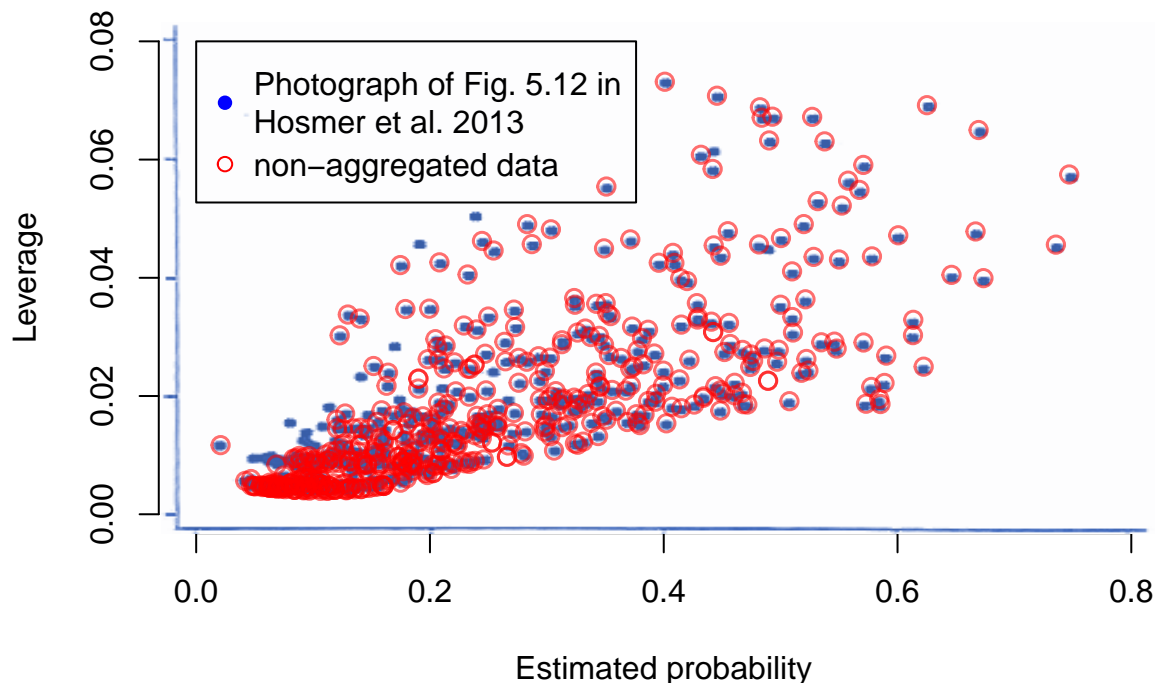


### Cook's distance based on package epiR vs. published figure 5.15



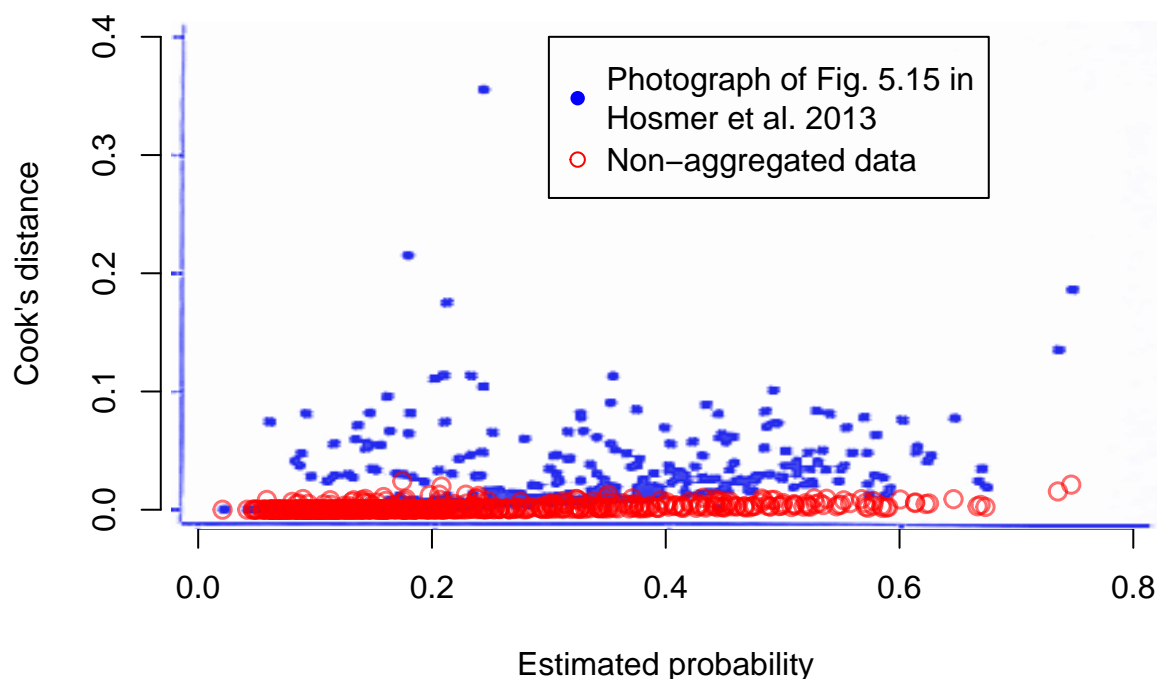
With leverages based on non-aggregated n-statistics instead of covariate patterns (m-statistics) the plot can be roughly reproduced, but there are still some points that do not match.

## Leverage based on non-aggregated data vs. published figure 5.12



But this remedy does not work for Cook's distance. The values based on n-statistics are completely different from those printed in the book and those calculated by epiR in magnitude and pattern.

## Cook's distance based non-aggregated vs. published figure 5.15

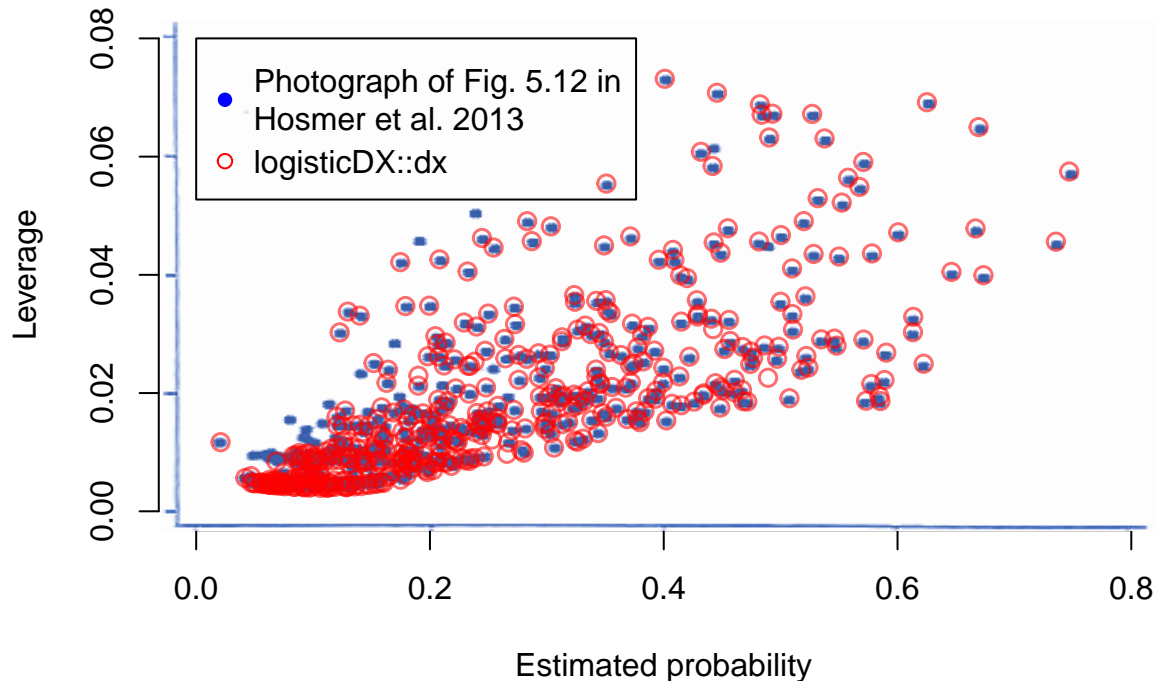


I tried out the newer package `logisticDx`<sup>5</sup> to calculate the diagnostic statistics. This package specializes on

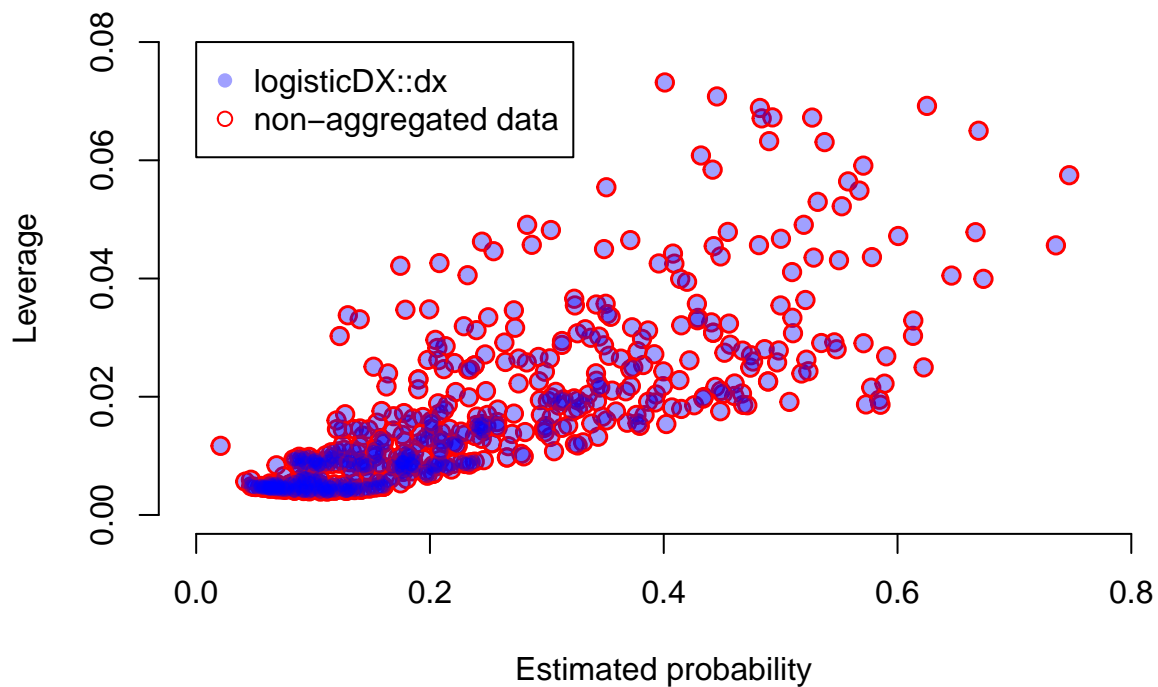
<sup>5</sup>Chris Dardis (2015). `LogisticDx`: Diagnostic Tests for Models with a Binomial Response. R package version 0.2. <https://github.com/cdardis/logisticDx>

diagnostic tests for regression models with binomial response and is explicitly based on the book of Hosmer et al. albeit in its second edition from 2000.

### Leverage based on package logisticDX vs. published figure 5.12



### Leverage based on logisticDX::dx vs. original data



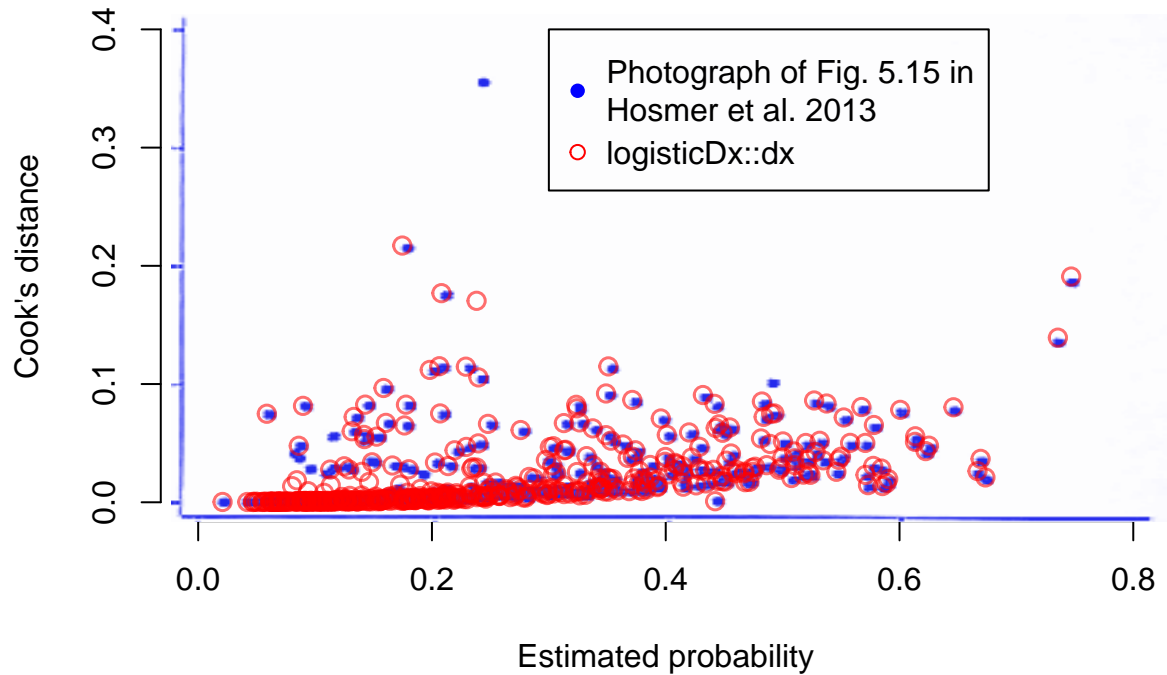
You can see, that the leverages of `logisticDX::dx` are exactly the same as those calculated from non-aggregated data (n-statistics). We have seen before that these are matching figure 5.12 quite well but not with all the

[//CRAN.R-project.org/package=LogisticDx](https://CRAN.R-project.org/package=LogisticDx)

points.

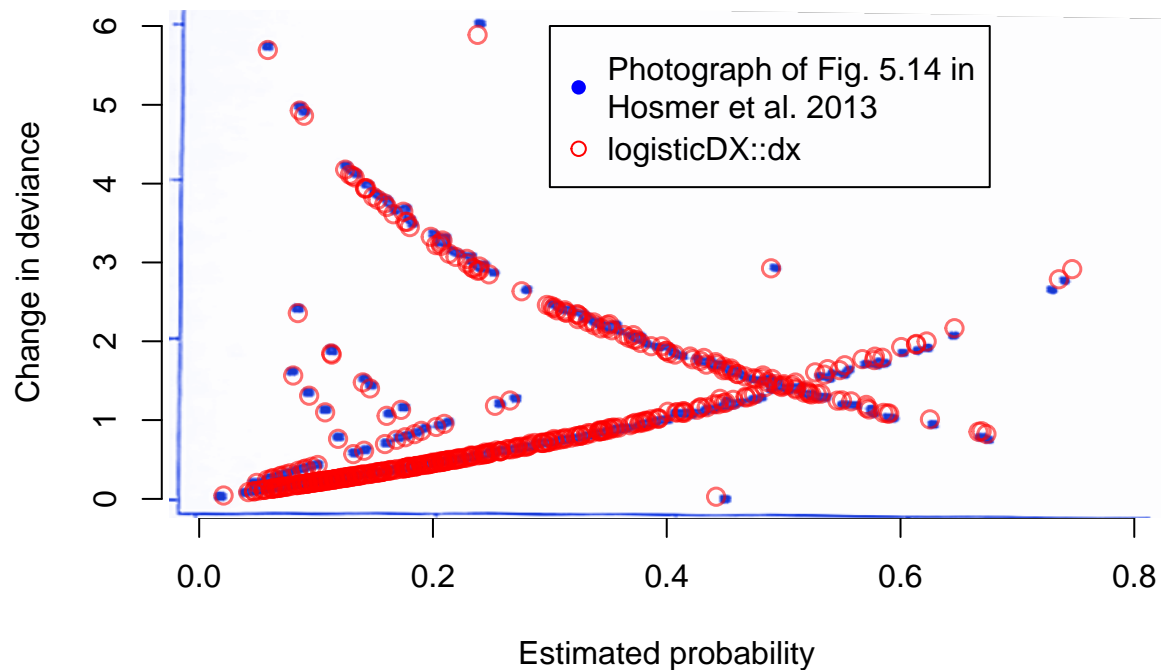
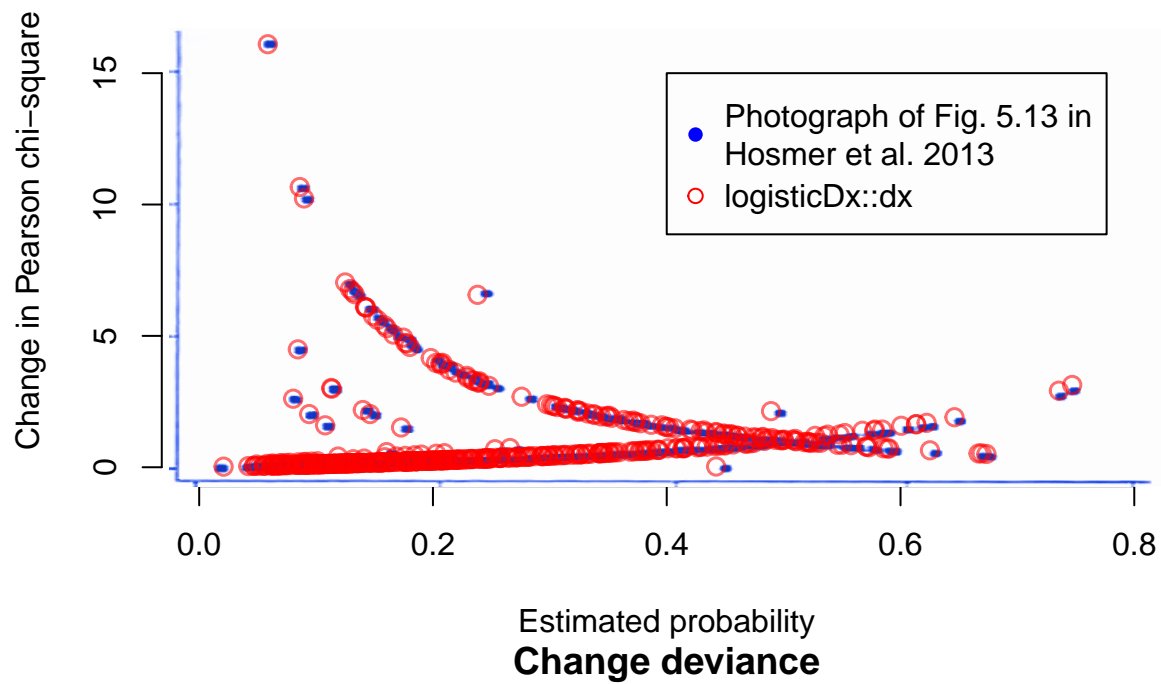
When they are calculated with `logisticDX::dx`, also the values of Cook's distance are fitting well with figure 5.15 despite of some single cases. I guess these are the same as those that do not fit in leverage.

### Cook's distance based on package `logisticDx` vs. published figure 5.



Last but not least here are the plots of change in Pearson chi-square and change in deviance. The packages `epiR` and `logisticDx` produce exactly the same results. I think that the mismatches in these graphs are due to my bad photographs: The book page was not completely plane but a little bit arched.

### Change in Pearson chi-square based on logisticDX::dx vs. published figure 5.13



I think that the mismatches in these graphs are due to my bad photographs: The book page was not completely plane but a little bit arched.