We address the need of considering a heavy–tail behaviour of contact degree distribution in social contact surveys

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

国立大学法人 長崎大学 NAGASAKI UNIVERSITY

Survey design matters for capturing heterogeneity in social contacts: from exponential to scale–free

Toshiaki R. Asakura[1,2,3,4], Sebastian Funk[1,2], Akira Endo[3,4,5],

[1] Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
[2] The Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK
[3] School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan
[4] Department of Infectious Disease Epidemiology and Dynamics, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan
[5] Saw Swee Hock School of Public Health, National University of Singapore, Singapore

Background

Substantial effort has been invested in collecting the number of social contacts per individual to characterise interactions between different groups. However, those studies usually focused only on the mean number of contacts, ignoring substantial variation therein, which might bias sample means.

Objectives

To explore the tail shape of the contact degree distribution across different social contact surveys and to identify key survey designs causing differences in tail behaviours.
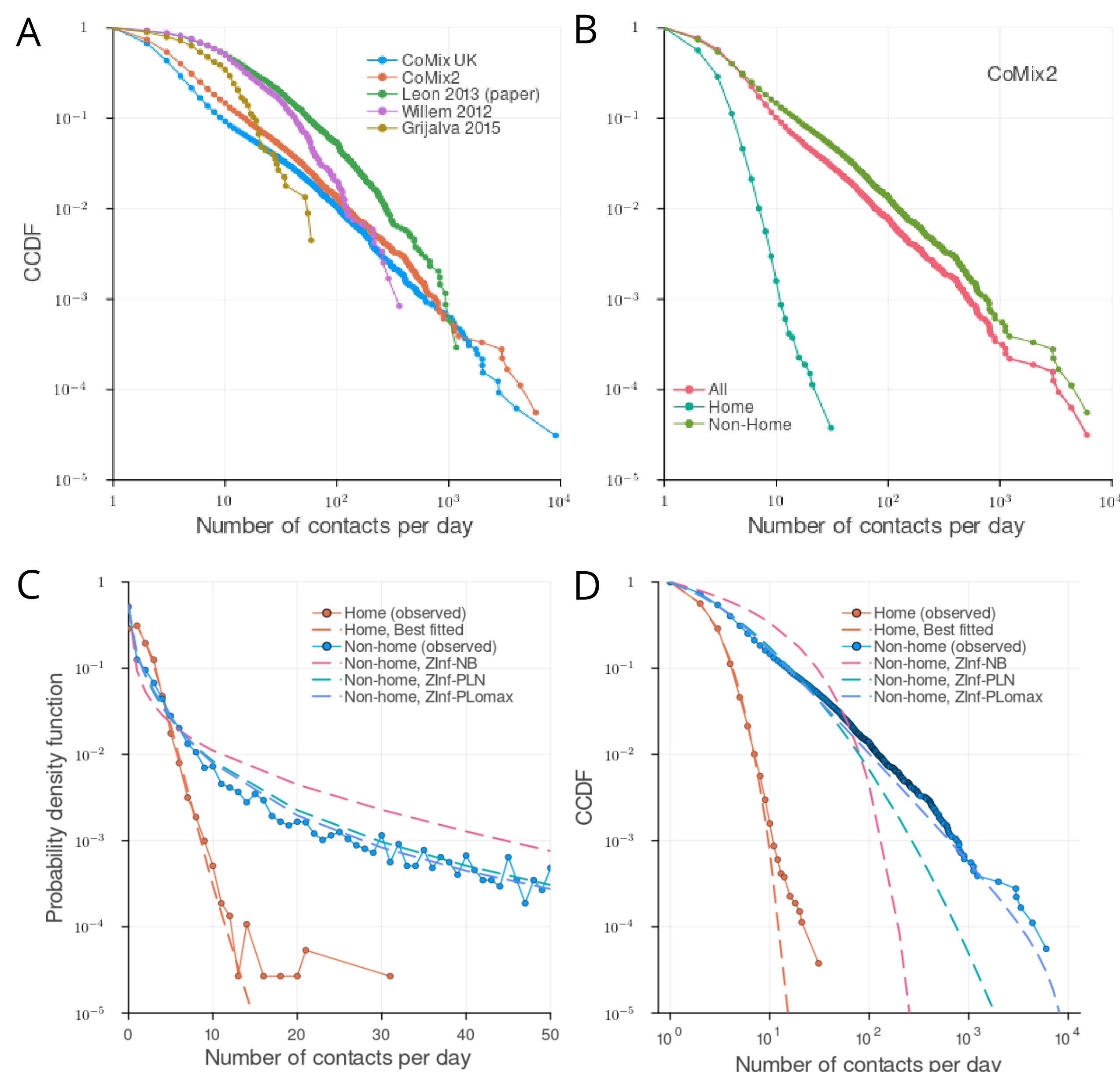
Data

· Collected 15 publicly available social contact surveys.
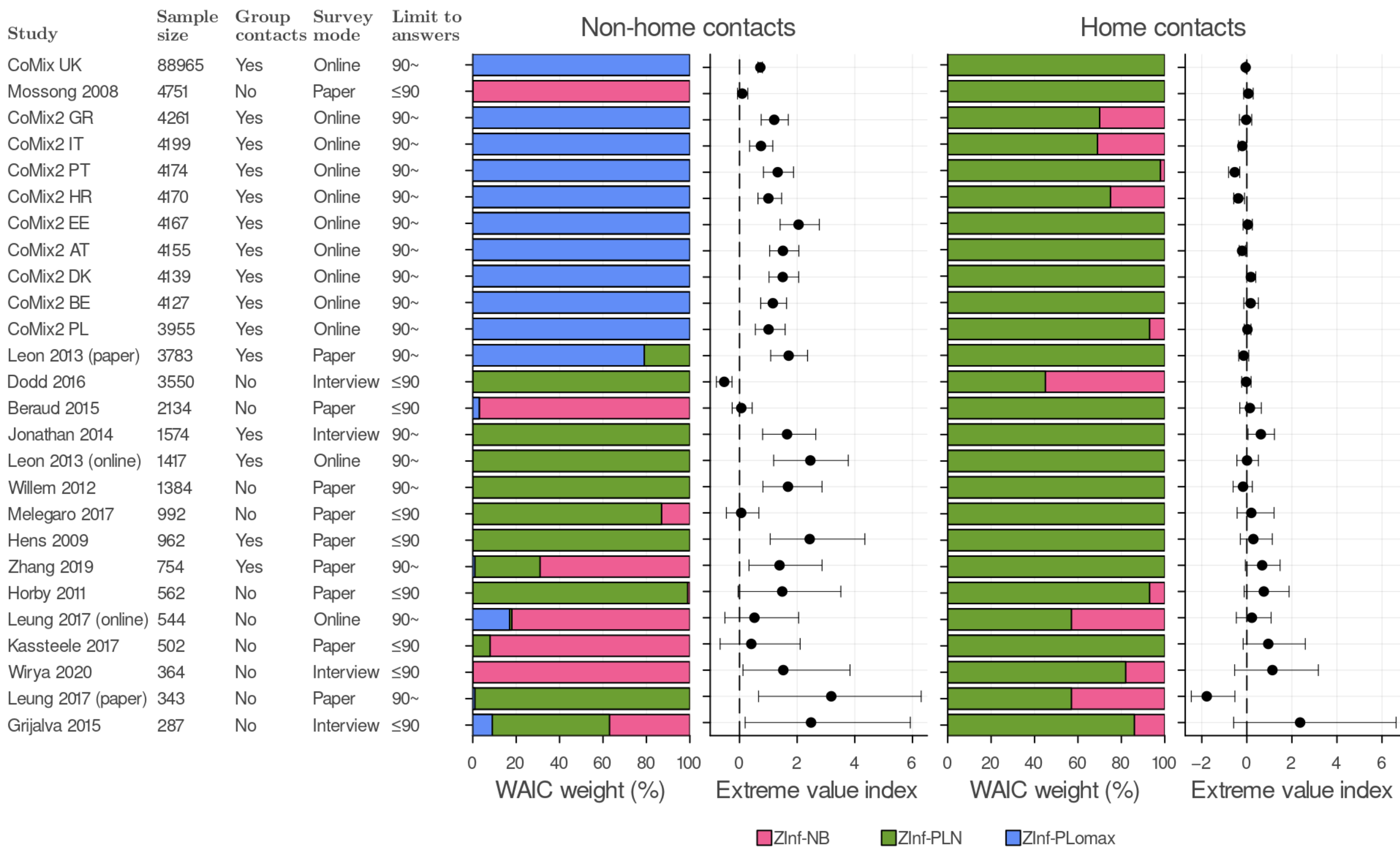· Included participants aged ≥18 years old.

Methods

1. Fitted three zero-inflated (Zinf) Poisson mixtures:
   - NB, negative binomial as an exponential family;
   - PLN, Poisson-lognormal as a heavy-tail distribution;
   - Plomax, Poisson-Lomax as a scale-free distribution.
2. Meta-regression on WAIC weights.
3. Assessed the biases by model misspecification when the true model is the Zinf-Plomax distribution.

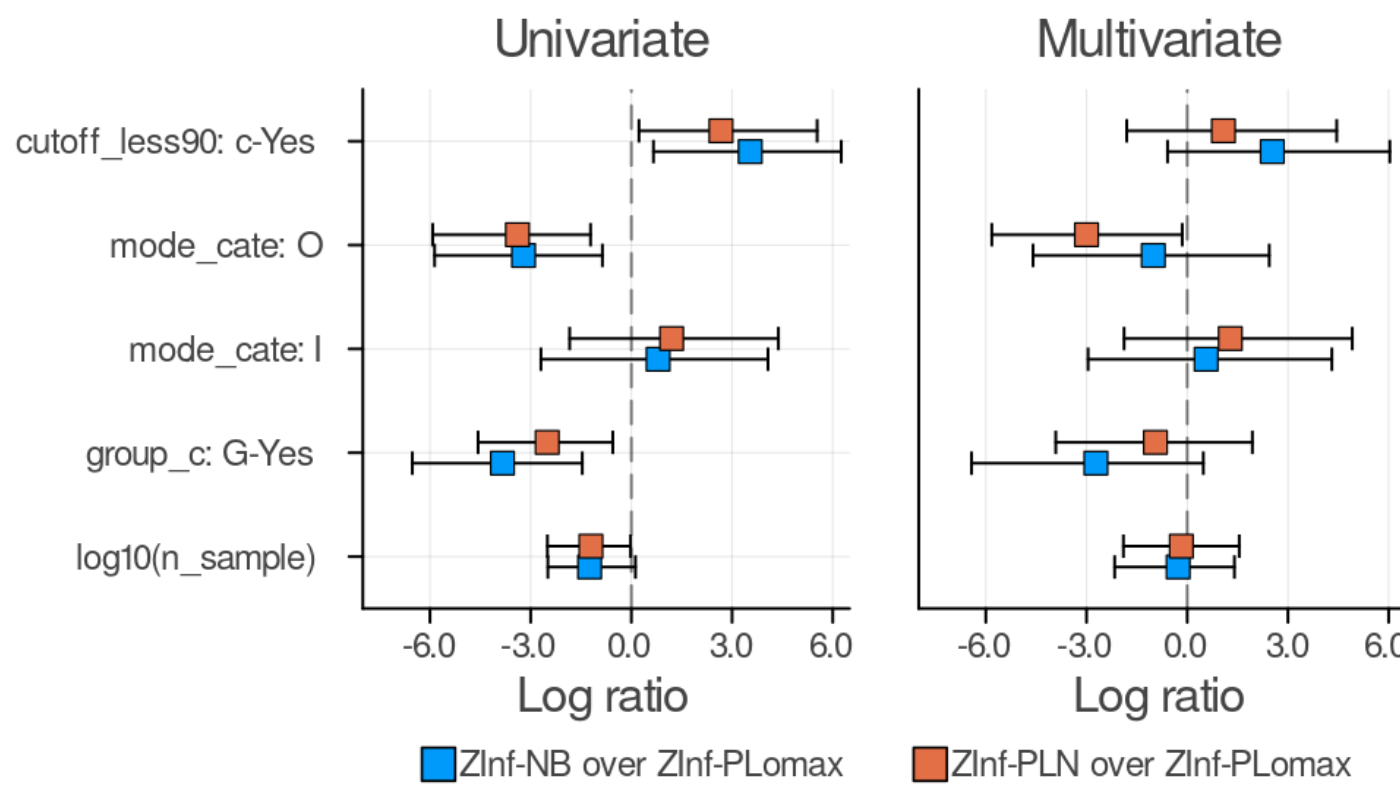### Separation of home and non-home contacts simplifies the shape of distributions

(A) Contact degree distributions across surveys, and (B) for CoMix2 separated by home and non-home. (C,D) Visualisation of fitting results for CoMix2. **NB poorly fitted to non-home contact degree distribution although frequently used.**

For non-home contacts, online surveys with group contacts favoured the ZInf-PLomax while the tail of some surveys are also scale-free

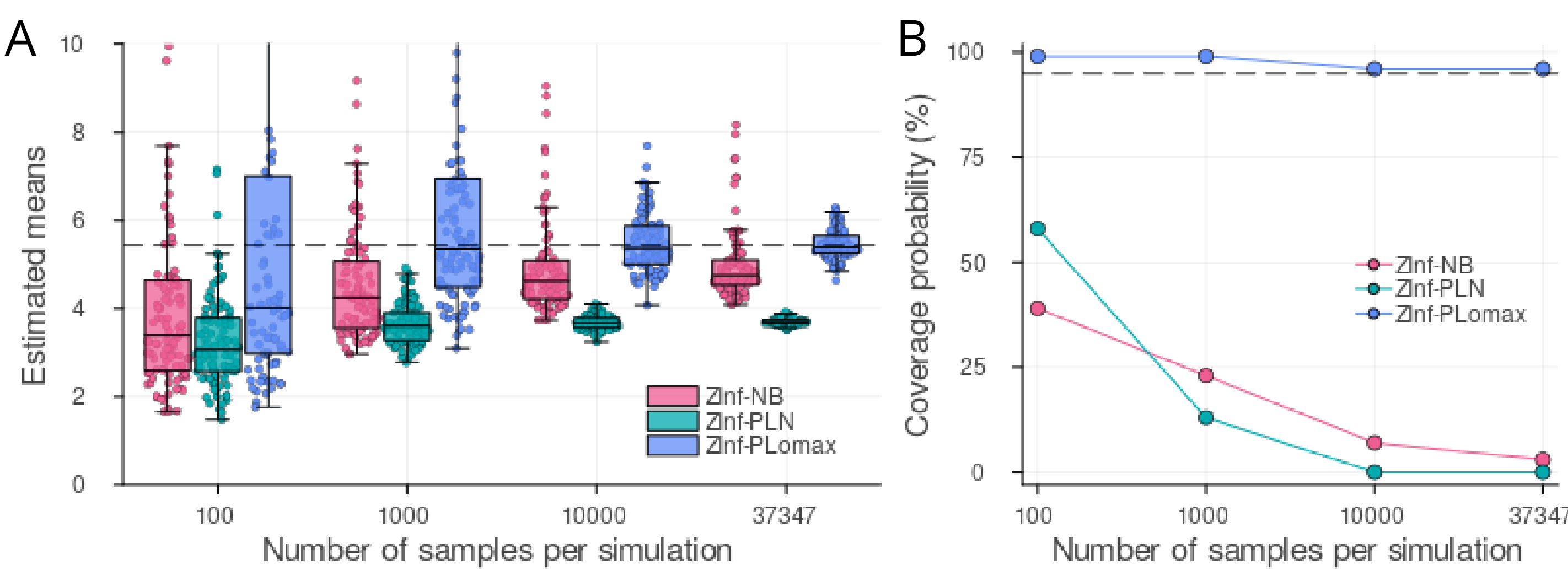| Study	Sample size	Group contacts	Survey mode	Limit to answers
CoMix UK	88965	Yes	Online	90~
Mossong 2008	4751	No	Paper	≤90
CoMix2 GR	4261	Yes	Online	90~
CoMix2 IT	4199	Yes	Online	90~
CoMix2 PT	4174	Yes	Online	90~
CoMix2 HR	4170	Yes	Online	90~
CoMix2 EE	4167	Yes	Online	90~
CoMix2 AT	4155	Yes	Online	90~
CoMix2 DK	4139	Yes	Online	90~
CoMix2 BE	4127	Yes	Online	90~
CoMix2 PL	3955	Yes	Online	90~
Leon 2013 (paper)	3783	Yes	Paper	90~
Dodd 2016	3550	No	Interview	≤90
Beraud 2015	2134	No	Paper	≤90
Jonathan 2014	1574	No	Interview	≤90
Leon 2013 (online)	1417	Yes	Online	90~
Willem 2012	1384	No	Paper	90~
Melegaro 2017	992	No	Paper	≤90
Hens 2009	962	Yes	Paper	≤90
Zhang 2019	754	Yes	Paper	90~
Horby 2011	562	No	Paper	≤90
Leung 2017 (online)	544	No	Online	90~
Kassteele 2017	502	No	Paper	≤90
Wirya 2020	364	No	Interview	≤90
Leung 2017 (paper)	343	No	Paper	≤90
Grijalva 2015	287	No	Interview	≤90

We calculated WAIC weights (strength of evidence for models in percentage measurement) and an extreme value index (>0 indicating those tails are scale-free).

Meta regression identifies online survey design and group contacts as a key for being scale-free

Fractional multinomial regression on WAIC weights was performed as a meta-regression. Negative log ratio means favouring Zinf-PLomax. Group contacts refer to reporting contacts without writing details of each contact.

Model misspecification by NB and PLN resulted in underestimation of means given PLomax is true model

(A) Estimated means for 100 sets of simulated data from Zinf-PLomax, fitted to non-home contact degree distribution of CoMix2. Horizontal dotted line represents the true mean. (B) Coverage probability (probability (i.e. proportion of instances where confidence interval includes the true mean) for each distribution.

Conclusions

· Separation of home and non-home contacts allows to describe the contact degree distribution with simple distributions (c.f. applying 4 parameter distributions to all contacts [1]).
· Scale-free property arises in online surveys with group contacts, and models considering heavy-tails should be used to estimate means.

[1] Danon, L., et al., (2013) Proc R Soc B. 280(1765), 20131037.