



# Multiple seed structure and disconnected networks in respondent-driven sampling

Jens Malmros<sup>\*†</sup>      Luis E.C. Rocha<sup>‡</sup>

March 2016

## Abstract

Respondent-driven sampling (RDS) is a link-tracing sampling method that is especially suitable for sampling hidden populations. RDS combines an efficient snowball-type sampling scheme with inferential procedures that yield unbiased population estimates under some assumptions about the sampling procedure and population structure. Several seed individuals are typically used to initiate RDS recruitment. However, standard RDS estimation theory assume that all sampled individuals originate from only one seed. We present an estimator, based on a random walk with teleportation, which accounts for the multiple seed structure of RDS. The new estimator can also be used on populations with disconnected social networks. We numerically evaluate our estimator by simulations on artificial and real networks. Our estimator outperforms previous estimators, especially when the proportion of seeds in the sample is large. We recommend our new estimator to be used in RDS studies, in particular when the number of seeds is large or the social network of the population is disconnected.

**Key words:** Respondent-driven sampling; Seeds; Disconnected network; Random walk with teleportation.

---

<sup>\*</sup>Department of Mathematics, Stockholm university, SE-106 91 Stockholm, Sweden

<sup>†</sup>Corresponding author: [jensm@math.su.se](mailto:jensm@math.su.se)

<sup>‡</sup>Karolinska Institutet, Stockholm, Sweden; Université de Namur, Belgium