Intra-Cluster Correlation Estimates for the Statistical Design of Trials in Household Homelessness

Michael Sanders and Dimitris Vallis

King's College London, UK
King's College London, UK

> **Abstract_** *Randomised controlled trials investigating homelessness have been utilised more frequently in the last few years to help evaluate, understand, and ultimately reduce the number of homeless households in the UK. During the designing stage of any trial, power calculations are used to help determine the required sample size and the minimum detectable effect of interest. These calculations, however, require a number of assumptions to be made regarding the structure and size of the data to be used. In order to support researchers looking to conduct trials in this field, we estimate intra-cluster-correlation-rate values that are essential in the design of such trials. We estimated ICCs for the incidence of households threatened/experiencing homelessness through time, categorised by region and by support needs. Intracluster correlation estimates ranged between 0.1 and 0.2 categorised at region level, reaching values up to 0.5 when further subcategorised by support needs. The existence of clustering of households threatened/experiencing homelessness emphasises the need for the provision of ICC values to researchers in order to facilitate the successful implementation of future randomised controlled trials.*

> **Keywords_** *RCT, power calculations, ICC, household homelessness*

Introduction

Situations such as being a young parent, having a history of abuse, or a history of mental health problems are all important determinants of homelessness. A number of trials have investigated these, as well as other related factors, in an attempt to better understand and ultimately help toward reducing the incidence of homelessness among families (Fowler et al., 2018; Vallesi et al., 2019; Yakubovich et al., 2022). However, the total number of trials in homelessness is small, with even fewer in the UK, despite a policy commitment to end homelessness, and the creation of the Centre for Homelessness Impact, a UK Government recognised "What Works Centre" dedicated to ending homelessness with evidence (Teixeira and Cartwright, 2020).

Canonical randomised controlled trials require randomisation at the individual level, with each individual randomised to receive either the intervention to be tested, or into a control group who will typically receive 'business as usual' support. In many, or perhaps most, cases, randomisation at that level may not be possible due to factors such as within-group non-independence of outcomes or the risk of intervention contamination. This is also true for people experiencing homelessness, with particular districts exhibiting a higher incidence as a result of area level factors. For example, areas with a higher level of unemployment, crime rate, and poverty will often be correlated with higher numbers of households experiencing homelessness (Fargo et al., 2013; Mabhala et al., 2020). Current UK based studies that aim to reduce homelessness, including Sanders and Picker (2023a), Sanders and Picker (2023b), and EDIT (2022), all rely, to some extent, on assignment of interventions at the level of a cluster, typically a geographical unit.

The effects of this clustering should be considered when deciding what the necessary sample size should be for a trial. Higher homogeneity of homelessness within each cluster would in turn imply the need for a higher sample size to achieve the required power.

Unlike quasi-experimental approaches or other designs that make use of secondary data, in which researchers are given a sample, designing a trial allows for a higher degree of freedom in calculating what the required sample size needs to be, given ethical or practical considerations. This involves determining how large a study must be relative to the desired minimum detectable effect size, reducing the chances of type II errors, increasing efficiency, and promoting ethical planning. These calculations are commonly referred to as power, or sample size calculations, and are one of the most important components in the design stage of an RCT.

Given that these calculations can make the difference between a robust trial and one from which nothing can be learned, it is integral that they are conducted well, and using accurate information. This is particularly challenging in the case of cluster randomised trials, in which the effect of clustering must often be guessed or approximated without adequate information. In this paper, we seek to support researchers in conducting these calculations by providing information on the statistical values to be used.

Materials and Methods

During the designing stage of a cluster randomised trial, there are certain parameters that play an important role in the size of the required sample. One of the most important ones is the intra-cluster correlation or ICC. This parameter is derived as a ratio of the relative magnitude of within- and between cluster variances of the outcome of interest. A large amount of within cluster homogeneity, or low variance, suggests that outcomes of subjects nested in clusters are very alike, reducing the amount of information provided from each cluster. This in turn would suggest the need of a higher sample size to achieve the required power. Contrarily, a study in which there is substantial variation within clusters, the per subject information available is higher.

The ratio of the sample size needed in an individually randomised trial to a cluster randomised trial is given by the 'design effect', which Kerry and Bland (1998), define as;

$$D = 1 + (m-1)\rho$$

Where D is the design effect, m is the number of observations in the average cluster, and ρ is the ICC.

As an example, assuming we are interested in detecting an effect of 0.3 standard deviations, with a probability of 80%. Our average cluster size is 50, and our ρ is 0.07. Using rudimentary power calculations conducted in the statistical software R with package 'pwr', we find that in an individually randomised trial, we will require 175 observations per arm. Using this we can then calculate our design effect as:

$$D = 1 + (50-1)0.07 = 4.4$$

Applying our design effect of 4.4 to our 175 observations implies we would need roughly 770 observations per arm in our cluster randomised trial to achieve the desired level of power. Assuming a higher ICC of 0.1 would in turn raise the required sample size per arm by approximately 250 observations, to a total of more than 1 000.

From this example, it is salient that clustering makes a substantial difference to the required sample, and that this difference depends heavily on the ICC. Overestimating the ICC will lead to unnecessarily large trials, leading to higher costs or barriers to conducting the trial due to lack of resources. Similarly, underestimating the ICC will lead to 'underpowered' trials suggesting a higher probability of a type II error (false negative).

Notwithstanding the importance of the ICC, many researchers resort to using rules of thumb or educated guesses during the designing stage of the trial, which is often their only option in the absence of empirical evidence. In order to help trialists in designing of trials in homelessness, we produce results of ICCs for the incidence of households threatened by or experiencing homelessness through time, categorised by region, as well ICCs for the number of households experiencing homelessness categorised by support needs.

The dataset used was of the detailed local authority level homelessness tables (DLUHC, 2023), which provide, among other statistics, the number of households threatened with homelessness and the number of households owed a homelessness duty. For the latter, numbers are also categorised by number of households owed a duty by the following support needs:

- Young person Aged 16-17 years

- Aged 18-25 years requiring support to manage independently

- Young parent requiring support to manage independently

- Care leaver aged 18-20 years

- Care leaver aged 21+ years

- Physical ill health and disability

- History of mental health problems

- Learning disability

- At risk of / has experienced sexual abuse / exploitation

- At risk of / has experienced domestic abuse

- At risk of / has experienced abuse (non-domestic abuse)

- Drug dependency needs

- Alcohol dependency needs

- Offending history

- History of repeat homelessness

- History of rough sleeping

- Former asylum seeker

- Old age

- Served in HM Forces

- Access to education, employment, or training

In order to estimate ICCs of homelessness at local authority level, we use the frequency of homeless households as a basis to generate pseudo-individualised units within each local authority. This allows us to estimate clustering effects on household-level data nested within local authorities without the risk of identifying households.

The model estimated is a binomial generalised linear mixed effects model of the form:

$$Prob(y_{ij} = 1) = \Lambda(\beta_{00} + u_{0j}) \text{ with } u_{0j} \, N(0, \sigma_u^2)$$

Where *Prob(y_{ij}=1)* is the conditional probability that the outcome variable equals one for household i in district j, and $\Lambda(x)$ is the logistic 'link' function $\frac{e^x}{1+e^x}$. Using this model, the ICC can then calculated using the following formula:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \frac{\pi^2}{3}}$$

Where σ_u^2 is the (level-2) random intercept variance and $\frac{\pi^2}{3}$ the (level-1) residual variance of the logistic distribution.

Results

Estimated ICCs are shown in the tables below, calculated for each region by year for households threatened with homelessness (Table 1) and households experiencing homelessness (Table 2).

Table 1. ICCs by region and year of households threatened with homelessness

	2022-2021	2021-2020	2020-2019	2019-2018
All	0.090	0.134	0.084	0.100
East Midlands	0.084	0.089	0.070	0.090
East of England	0.076	0.104	0.079	0.091
London	0.127	0.131	0.096	0.096
North East	0.077	0.128	0.134	0.104
North West	0.094	0.170	0.111	0.080
South East	0.062	0.134	0.058	0.112
South West	0.075	0.189	0.064	0.094
West Midlands	0.108	0.112	0.071	0.142
Yorkshire and The Humber	0.077	0.108	0.073	0.047

Table 2. ICCs by region and year of households experiencing homelessness

	2022-2021	2021-2020	2020-2019	2019-2018
All	0.120	0.105	0.102	0.096
East Midlands	0.097	0.119	0.118	0.096
East of England	0.070	0.077	0.082	0.069
London	0.274	0.115	0.077	0.076
North East	0.070	0.085	0.072	0.065
North West	0.091	0.118	0.119	0.120
South East	0.111	0.089	0.105	0.098
South West	0.116	0.123	0.103	0.098
West Midlands	0.149	0.102	0.091	0.095
Yorkshire and The Humber	0.093	0.101	0.097	0.102

Overall, estimates range around an ICC of 0.1 to 0.2, with the largest values observed for the region of London followed by the North and South West, suggesting stronger clustering effects compared to other regions. There is a faint indication of some increase in the estimates as we move toward the more recent dates, with 2020-21 year exhibiting the highest estimates over all regions for households threatened with or experiencing homelessness.

Tables 3 and 4 contain ICCs by support needs and region, estimated for the most recent years of 2022-2021 and 2021-2020.

Table 3. Household homelessness ICCs by support needs and region (2022-21)

	All	East Midlands	East of England	London	North East	North West	South East	South West	West Midlands	Yorkshire and The Humber
Aged 16-17 years	0.231	0.246	0.156	0.286	0.191	0.138	0.211	0.214	0.288	0.204
Aged 18-25 years requiring support	0.161	0.126	0.134	0.254	0.091	0.188	0.097	0.152	0.131	0.167
Young parent requiring support	0.187	0.160	0.142	0.174	0.127	0.185	0.135	0.227	0.234	0.312
Care leaver 18-20 years	0.160	0.103	0.233	0.188	0.133	0.064	0.112	0.202	0.151	0.140
Care leaver aged 21+ years	0.221	0.136	0.127	0.253	0.387	0.168	0.236	0.218	0.112	0.231
Physical ill health/disability	0.141	0.106	0.078	0.182	0.174	0.099	0.149	0.146	0.155	0.105
Mental health problems	0.149	0.112	0.079	0.158	0.184	0.096	0.183	0.122	0.178	0.125
Learning disability	0.206	0.174	0.162	0.168	0.241	0.156	0.190	0.306	0.153	0.244
sexual abuse / exploitation	0.240	0.167	0.208	0.237	0.435	0.194	0.214	0.241	0.173	0.226
domestic abuse	0.122	0.073	0.082	0.144	0.147	0.097	0.148	0.111	0.118	0.107
abuse (non-domestic abuse)	0.208	0.206	0.202	0.248	0.235	0.172	0.187	0.184	0.153	0.212
Drug depend-ency needs	0.178	0.220	0.116	0.165	0.214	0.122	0.139	0.168	0.089	0.190
Alcohol dependency needs	0.156	0.136	0.138	0.151	0.173	0.134	0.103	0.180	0.094	0.142
Offending history	0.226	0.228	0.132	0.225	0.259	0.190	0.176	0.214	0.181	0.325
Repeat homelessness	0.301	0.286	0.230	0.229	0.352	0.349	0.264	0.363	0.199	0.295
History of rough sleeping	0.269	0.264	0.184	0.249	0.378	0.232	0.203	0.313	0.253	0.316
Former asylum seeker	0.305	0.293	0.191	0.290	0.393	0.161	0.269	0.209	0.078	0.443
Old age	0.124	0.058	0.102	0.165	0.122	0.171	0.077	0.152	0.157	0.086
Served in HM Forces	0.222	0.109	0.163	0.305	0.219	0.209	0.095	0.211	0.192	0.234
Access to education, employment or training	0.374	0.464	0.376	0.381	0.338	0.331	0.330	0.424	0.193	0.318

Table 4. Household homelessness ICCs by support needs and region (2021-20)

	All	East Midlands	East of England	London	North East	North West	South East	South West	West Midlands	Yorkshire and The Humber
Aged 16-17 years	0.247	0.177	0.217	0.286	0.255	0.260	0.115	0.247	0.172	0.226
Aged 18-25 years requiring support	0.167	0.256	0.142	0.150	0.165	0.181	0.078	0.133	0.150	0.254
Young parent requiring support	0.200	0.153	0.192	0.250	0.198	0.288	0.103	0.194	0.074	0.297
Care leaver 18-20 years	0.137	0.197	0.138	0.139	0.141	0.086	0.115	0.143	0.067	0.046
Care leaver aged 21+ years	0.182	0.127	0.168	0.134	0.165	0.132	0.185	0.219	0.142	0.174
Physical ill health/disability	0.144	0.137	0.095	0.131	0.181	0.145	0.134	0.131	0.147	0.105
Mental health problems	0.149	0.140	0.117	0.105	0.170	0.221	0.117	0.116	0.165	0.131
Learning disability	0.207	0.168	0.187	0.165	0.206	0.231	0.185	0.202	0.215	0.227
sexual abuse / exploitation	0.250	0.233	0.214	0.204	0.250	0.303	0.243	0.320	0.117	0.208
domestic abuse	0.150	0.088	0.133	0.115	0.096	0.149	0.201	0.137	0.128	0.100
abuse (non-domestic abuse)	0.231	0.246	0.233	0.219	0.277	0.189	0.203	0.257	0.213	0.211
Drug dependency needs	0.197	0.235	0.150	0.195	0.178	0.150	0.211	0.153	0.126	0.177
Alcohol dependency needs	0.151	0.187	0.103	0.140	0.129	0.089	0.137	0.188	0.119	0.117
Offending history	0.212	0.224	0.159	0.254	0.238	0.193	0.178	0.212	0.138	0.201
Repeat homelessness	0.283	0.282	0.243	0.251	0.421	0.245	0.235	0.394	0.160	0.333
History of rough sleeping	0.276	0.325	0.257	0.239	0.354	0.344	0.234	0.252	0.177	0.307
Former asylum seeker	0.316	0.376	0.208	0.199	0.322	0.198	0.329	0.303	0.280	0.279
Old age	0.145	0.113	0.160	0.135	0.103	0.126	0.093	0.165	0.221	0.060
Served in HM Forces	0.237	0.146	0.172	0.307	0.228	0.118	0.211	0.204	0.148	0.162
Access to education, employment or training	0.351	0.391	0.302	0.366	0.538	0.386	0.223	0.366	0.179	0.567

Results of ICCs by support needs indicate a higher level of heterogeneity among estimates among regions for both years. Values for all categories and regions, for the period of 2021-20, range between 0.05 and 0.57, while for the period of 2022-21, estimated ICCs are slightly lower, ranging between 0.06 and 0.46. This observation also reflects the observed difference in estimated ICCs through time. When comparing estimates between tables 3 and 4, it is evident that values have generally fallen as we move from the previous period to the next. Contrary to the above, the region of London is the most notable exception to this rule, where values have instead mostly risen. Contrary to the variation observed for regions through time, when considering aggregate estimates for all of England (see column "All"), values seem to remain relatively stable.

Higher-than-average ICC values were estimated for the following categories:

- Care leavers,

- People at risk of/experienced abuse,

- People with a history of repeat homelessness/rough sleeping,

- Former asylum seekers, and

- People in need of education, employment, or training.

The reasons for the higher observed ICCs for these categories could be related to area-level effects, which these groups may be particularly sensitive to. We would thus expect a larger design effect as a result of the higher estimated ICCs in these cases, and thus larger sample size requirements.

As these estimates are specific to England, ICC values of the categories explored may be characterised by some variation in other countries, particularly given the importance of country-specific policies targeted at vulnerable groups experiencing homelessness; conservatism is therefore warranted when considering these values for trials outside of England. However, given the literature's severe lack in availability of ICC estimates for household homelessness and homelessness in general, we believe that the aggregate estimates provided in this paper could serve, at the very least, as a useful starting point for researchers when considering the required power for trials in other countries.

Discussion

This paper introduced calculations of ICC parameters that can aid in the implementation of randomised controlled trials targeted at household homelessness. This was performed by creating pseudo-individualised respondents within each local authority, allowing for the investigation of clustering of outcomes. Results suggested low-to-moderate levels of clustering with ICCs ranging from 0.05 to 0.5. Predictably, certain categories exhibited higher ICC values, particularly for vulnerable groups such as people at risk of/experienced abuse, people with a history of homelessness, and former asylum seekers.

Disclosure Statement

No conflict of interest to report.

❯ References

Department for Levelling Up, Housing and Communities (2023) *Tables on Homelessness.* [online] Accessed at: https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness.

EDIT – The Policy Institute (2022) *Additional Financial Assistance for People with Experience of Homelessness Trial*. [online] Accessed at: https://osf.io/s8f3t.

Fargo, J.D., Munley, E.A., Byrne, T.H., Montgomery, A.E., and Culhane, D.P. (2013) Community-Level Characteristics Associated with Variation in Rates of Homelessness Among Families and Single Adults, *American Journal of Public Health* 103(Suppl2) pp.340-347.

Fowler, J.P., Brown, S.D., Schoeny, M., and Chung, S. (2018) Homelessness in the Child Welfare System: A Randomized Controlled Trial to Assess the Impact of Housing Subsidies on Foster Care Placements and Costs, *Child Abuse & Neglect* 83(1) pp.52-61.

Kerry, S.M. and Bland, J.M. (1998) Statistics Notes: Sample Size in Cluster Randomisation, *Bmj* 316(7130) p.549.

Mabhala, M., Esealuka, W.A., Nwufo, A.N., Enyinna, C., Mabhala, C.N., Udechukwu, T., Reid, J., and Yohannes, A. (2020) Homelessness Is Socially Created: Cluster Analysis of Social Determinants of Homelessness (SODH) in North West England in 2020, *International Journal of Environmental Research and Public Health*18(6) pp.3066.

Sanders, M. and Picker, V. (2023a) Investigating Interventions to Reduce Homelessness Among Care Leavers in Greater Manchester. [online] Accessed at: https://osf.io/3vcsz.

Sanders, M. and Picker, V. (2023b) Quasi-Experimental Evaluations of Interventions to Reduce Homelessness Among Care Leavers. [online] Accessed at: https://osf.io/pm65e

Teixeira, L. and Cartwright, J. (2020) *Using Evidence to End Homelessness* (Bristol: Policy Press).

Vallesi, S., Flatau, P., Thielking, M., Mackelprang, J.L., Taylor, K.M., La Sala, L., Spiers, J., Wood, L., Martin, K., Kragt, D., Lester, L., Whittaker, E., and Courtney, R.J. (2019) A Mixed Methods Randomised Control Trial to Evaluate the Effectiveness of the Journey to Social Inclusion – Phase 2 Intervention for Chronically Homeless Adults: Study Protocol, *BMC Public Health* 19(1) p.334.

Yakubovich, R.A., Bartsch, A., Metheny, N., Gesink, D., and O'Campo, P. (2022) Housing Interventions for Women Experiencing Intimate Partner Violence: A Systematic Review, *The Lancet Public Health* 7(1) pp.23-35.