Correspondence

Self weighing & non-probability samples

Sir,

I read with interest article on high prevalence of cardiovascular risk factors in Asian Indians; the Chandigarh Urban Diabetes Study (CUDS) published recently¹. The authors need to be appreciated for their effort in conducting a field based study on the prevalence of risk factors. However, I have some concerns with the methodology used and the statistics applied thereof.

The authors state that this was a cross-sectional population survey using multi-stage cluster randomized sampling conducted in Chandigarh, north India, involving 2227 subjects. They further state that two sectors from each of the three zones were selected by simple random sampling. The first house was selected from within each selected sector by simple random sampling. Starting from that house, all the eligible people ≥ 20 yr of age were screened from the consecutive houses till a sample size of at least 375 was reached in that sector¹.

Although there is no clear cut methodological definition for multi-stage cluster randomized sampling, but the design that has been used by the authors is multi-stage cluster sampling not the multi-stage cluster randomized sampling. In multi-stage cluster sampling units are the groups of elements except for the last stage of sampling as done by the authors. The authors state that the probability of selecting each study subject from the total population was not uniform - *i.e.* the design was not a 'self-weighing' one. They further state that the inverse of the probability of selection was, therefore, employed as a weight for that subject.

Now if the design was strictly multi-stage cluster randomized sampling, then the probability of selecting each study subject must be uniform and not non-

uniform as stated by the authors. In a survey of this kind, we are concerned only with the probability sampling in which each element of a population has a known (non zero) probability of being included in the sample. This is the basis for applying statistical theory in the derivation of the properties of the survey estimators for a given design. Second, if a sample is to be drawn from a population, it is necessary to be able to construct a sampling frame that lists suitable sampling units that encompass all elements of the population. If it is not feasible or is impractical to list all population elements, some clusters of elements can be used as sampling units. For example, it is impractical to construct a list of all individuals in Chandigarh, but we can select the sample in several stages. In the first stage, zones are randomly sampled; in the second stage, sectors within the selected zones are sampled; in the third stage, street/ blocks are sampled within the selected sectors. Then, in the final stage of selection, a list of individual households is needed only for the selected blocks from which eligible individuals could be selected randomly. This multi-stage design satisfies the requirement that all population elements have a known non zero probability of being selected.

Sunil Kumar Raina

Department of Community Medicine Dr. RP Government Medical College, Tanda, Kangra 176 001, India ojasrainasunil@yahoo.co.in

Reference

 Walia R, Bhansali A, Ravikumar P, Bhadada SK, Shanmugaundar G, Dutta P, *et al.* High prevalence of cardiovascular risk factors in Asian Indians: A community survey- Chandigarh Urban Diabetes Study (CUDS). *Indian J Med Res* 2014; *139*: 252-9.