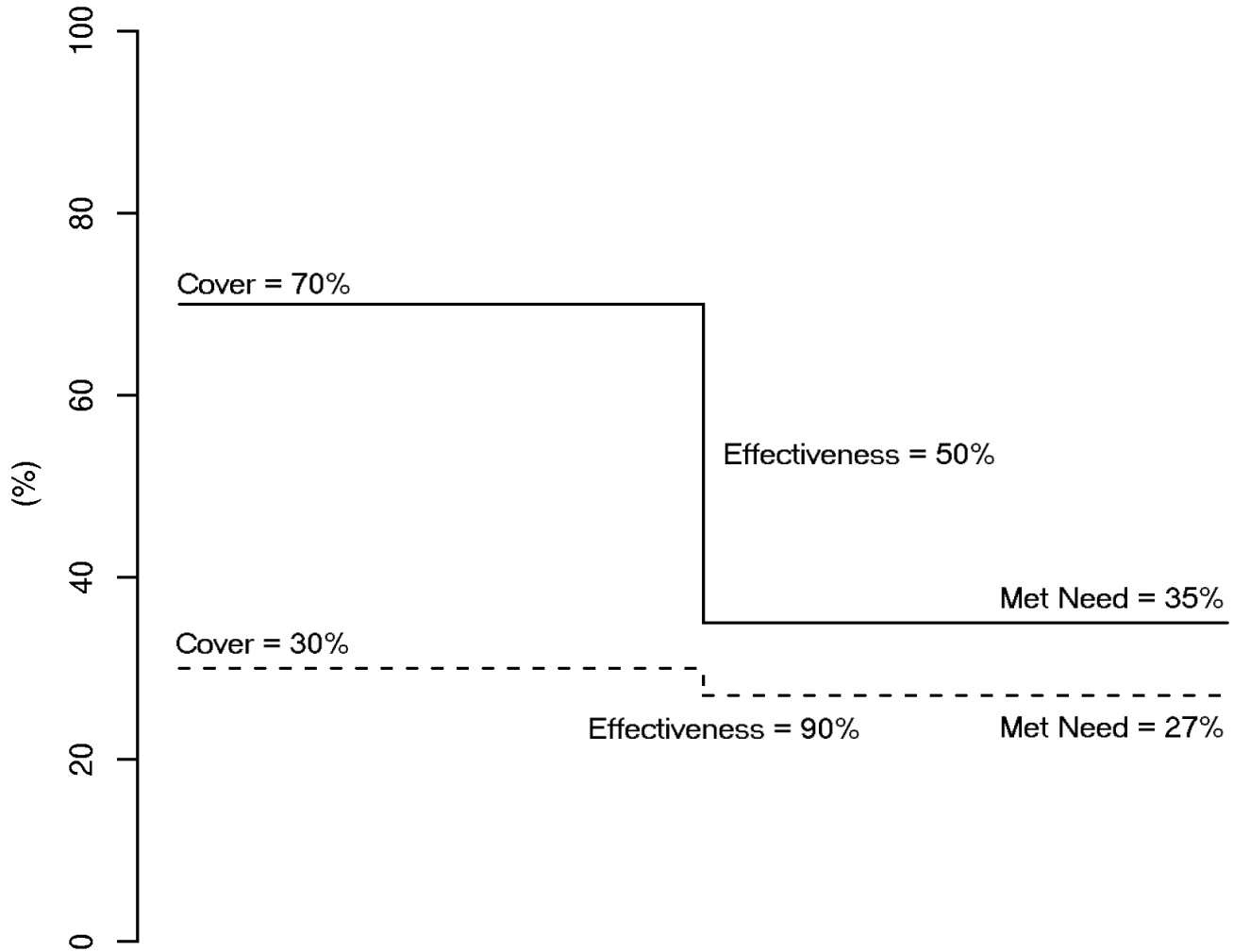
Coverage Survey Techniques

Mark Myatt, Institute of Ophthalmology

Structure of this presentation

1. Review of the relationship between cover, in-program effectiveness, and met need.
2. Review of current methods for estimating coverage and problems associated with them.
3. Outline of the method used to evaluate coverage in the Malawi CTC study and subsequent CTC implementations.
4. Results from the Malawi CTC study.

Cover / in-program effectiveness / met need



Current methods

- Two-stage cluster sampled nutritional anthropometry surveys are used to estimate coverage:
  - *Directly* using survey data.
  - *Indirectly* using survey data, program data, and population estimates.

Current surveys

- Two-stage cluster-sampling approach:
  - *Population proportional sampling* (PPS) in first stage (select clusters).
  - *Proximity sampling* in second stage (select households and children).
  - Assumes *homogeneity* of coverage.
  - Coverage surveys 'bolted-on' to nutrition surveys:
    - Sample size problems.

Population Proportional Sampling

- Bulk of data collected from the most populous areas / communities:
  - Some low population-density areas **not** sampled (potential for upward bias in coverage estimate).
- No guarantee of an even spatial sample:
  - Some areas usually unrepresented in sample.
- Relies on population estimates.

Proximity sampling

- Not EPSeM (i.e. unlikely to be representative) at the cluster level:
  - No estimation / comparison at cluster level.
- Even if an EPSeM sampling method is used:
  - Within-cluster sample size too small to estimate coverage within cluster:
    - n = 900, p = 5%, cases = 45, cases / cluster = 1.5

Real problems?

- These problems are **not** important if the homogeneity assumption is true:
  - Unlikely to be true of some centre-based programs or during start-up phase.
  - Difficult to test without a more expensive survey ... then a survey is not needed.

If the homogeneity assumption is untrue …

- Coverage is uneven and ...
  - It is useful to be able to identify:
    - Where coverage is *good*.
    - Where coverage is *poor*.
- But … current methods only provide a single wide-area coverage estimate.

Sample size (direct method)

- *Best case* example:
  - 30 x 30 design, n = c. 900
  - Assume:
    - Prevalence = 5%
    - Coverage = 50%
    - Design effect = 2.0
      - n = 45 cases (i.e. 5% of 900)
      - Estimate = 50% ± c. 20%

∴ Sample size is too small to estimate overall coverage with *reasonable precision*.

Denominator (indirect method)

- Coverage estimated as:

$$\frac{number\ of\ recipients\ attending\ program}{estimated\ prevalence\ of\ severe\ acute\ undernutriton \times estimated\ population} \times 100$$

- Unstable / unreliable denominator:
  - Prevalence estimate is *relatively* imprecise:
    - e.g. 5% ± 2% = 40% *relative* precision
  - Population estimate difficult to correct for displacement, migration, and high mortality in the target population.

Recycled data (indirect method)

- Usually applied when data (i.e. from a current survey) for the direct estimation is **not** available:
  - Initial assessment data:
    - Historic rather than current prevalence estimate.

Data needs

- Large data-entry overhead.
- Data management expertise:
  - Calculate nutritional indices.
  - Apply case-definitions.
  - Subsets of data.
- Multi-stage design requires specialist software to calculate confidence intervals.

Summary of problems

- Implausible homogeneity assumption.
- Uneven spatial sampling.
- Inadequate sample size (cluster / overall).
- Denominator problems (indirect method).
- Data-management overhead.
- Potential for inappropriate data-analysis and misreporting.

Attributes of a replacement

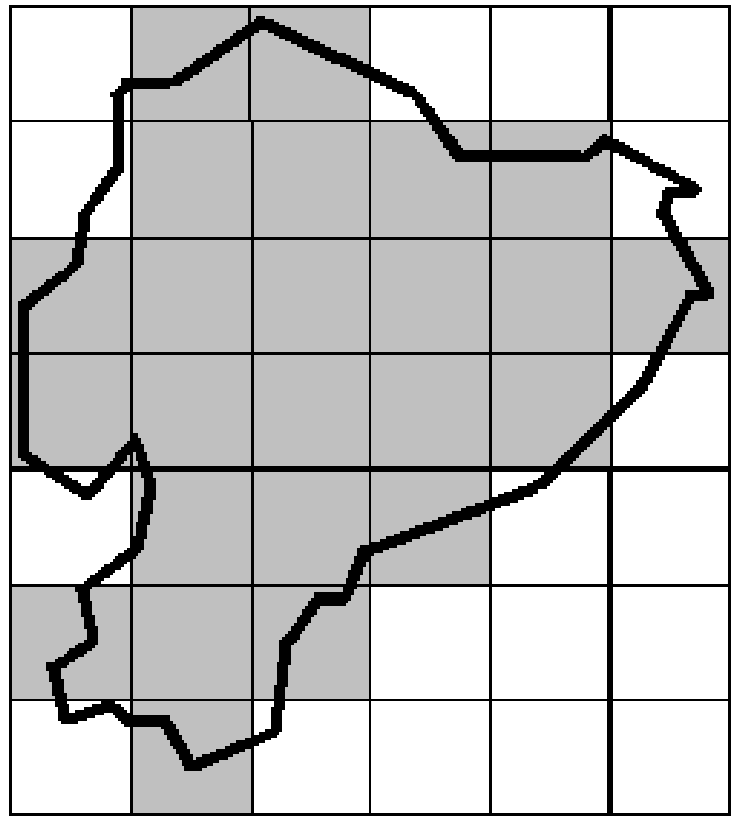
- Plausible homogeneity assumption.
- Even spatial sampling.
- Adequate within-community sample size:
  - To allow local estimation and comparison.
- Denominator from survey.
- Little data-entry and management.
- Simple data analysis.

Plausible homogeneity assumption

- Stratified design:
  - Many small (geographical) strata make the homogeneity assumption reasonably plausible:
    - Area of individual strata typically 1/30 of survey area:
      - Surveys presented here used 10km x 10km ($100km^2$) strata.

Even spatial sampling

- Strata defined using the *centric systematic area sampling* (CSAS) approach.
- Sample communities closest to the centre of each *quadrat*.
- c. 30 strata used.
- Exhaustive geographical sample.
- Even spatial sample.



Adequate Sample Size

- Probability sampling abandoned.
- Active case-finding adopted.
- Comparison of active case-finding against whole community screening:
  - Active case-finding returns an *exhaustive* sample:
    - Allowing small and wide area estimates / comparisons.
  - Active case-finding is rapid:
    - Six communities screened by one team in one day.

Denominator

- Derived from survey:
  - Number of cases found:
    - In each quadrat.
    - In survey area.
- Two levels of estimate available:
  - Quadrat.
  - Survey area.

Data needs

- One record per quadrat, typically:
  - Number screened.
  - Number in program.
  - Number of cases found.
  - Quadrat identifiers (x, y).
- Low data-entry overhead:
  - c. 150 data items vs. c. 6300 data items.
  - Can be done in a spreadsheet.

Simple data analysis

- Data may be treated as either coming from:
  - A simple random sample.
  - A stratified sample:
    - Analysis weighted by the populations of sampled communities in each quadrat.
- Both are easy to do in a spreadsheet.

An additional advantage

- More communities are sampled than in a 30 x 30 survey:
  - CSAS/CF >= 150 communities
  - 30 x 30 <= 30 communities
- Overall:
  - More communities = larger sampling fraction of communities = more representative sample.
- Within strata:
  - Six communities / every HH (vs. 30 HH)
    - large sampling fraction = more representative sample.

Disadvantages

- Comparatively expensive, an example from the Malawi CTC study:
  - CSAS/CF 123 person-days
  - 30 x 30 81 person-days
- Why?
  - CSAS/CF 1 quadrat / team / day
  - 30 x 30 2 clusters / team / day
- But … method could be integrated with program case-finding activities.

Comparison of methods

	30 x 30	CSAS/CF
Design	Cluster	Stratified
First-stage sampling	PPS	Centric systematic area
Second stage sampling	Proximity	Exhaustive (case-finding)
Communities sampled	30	151
Children measured	1025	1403
Staff required (person-days)	81	123
Cases found	10	136
Cases covered	1	29
Cover (%)[1]	10.0%	20.0%
95% confidence interval	0.2%, 44.5%	13.8%, 26.3%

[1]Point estimates (direct)

Summary

- CSAS/CF addresses the problems inherent in current methods but …
  - It is more expensive.
- However …
  - It could be integrated into program case-finding activities.
  - It could be adapted to estimate prevalence and coverage (therapeutic and supplementary) simultaneously.

Results

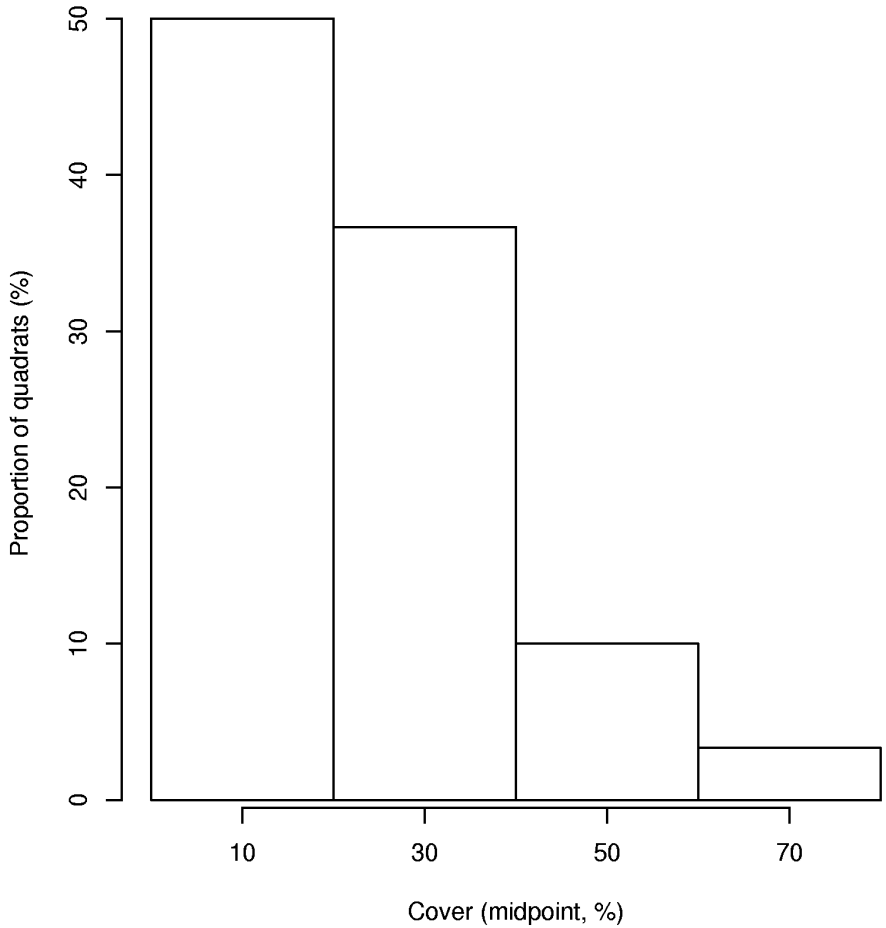
- Overall coverage estimate.
- Per-quadrat coverage estimates:
  - Range.
  - Map.

Overall coverage (TFC)

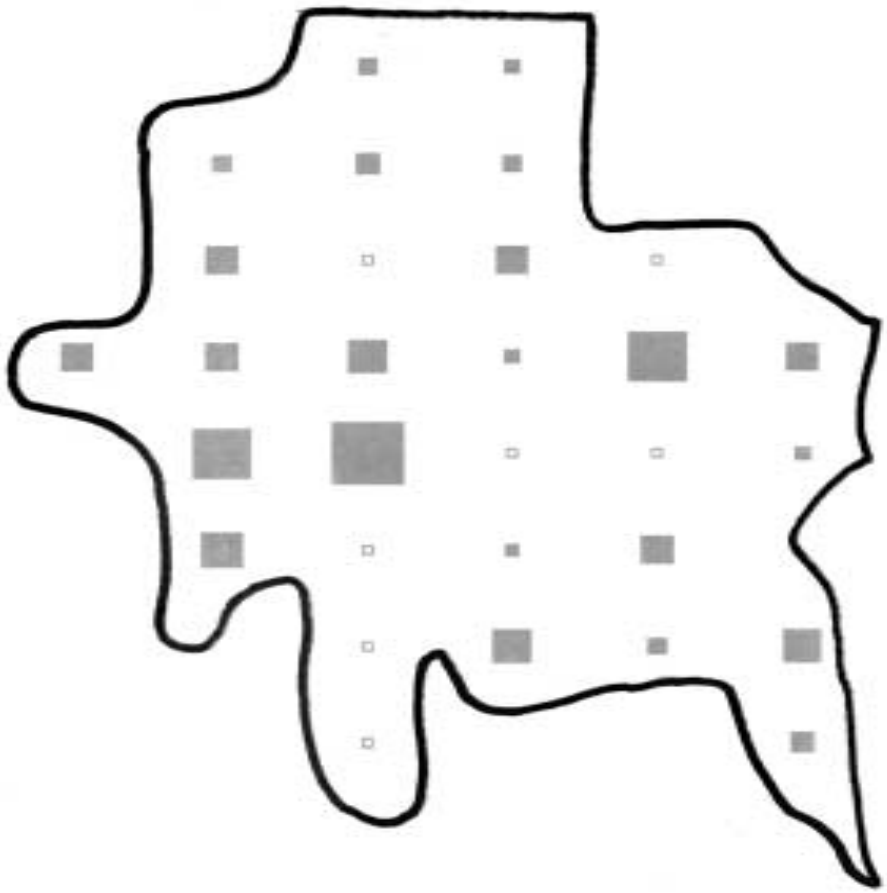
26.3%

(95% CI = 19.1%, 34.5%)

Range of coverage (TFC)



Map of coverage (TFC)



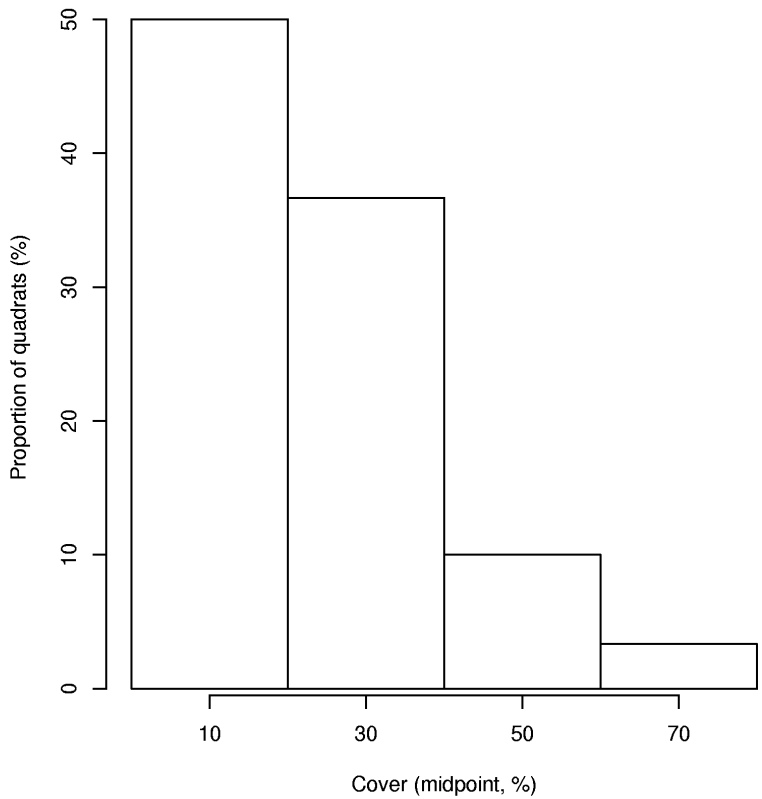
TFC vs. CTC

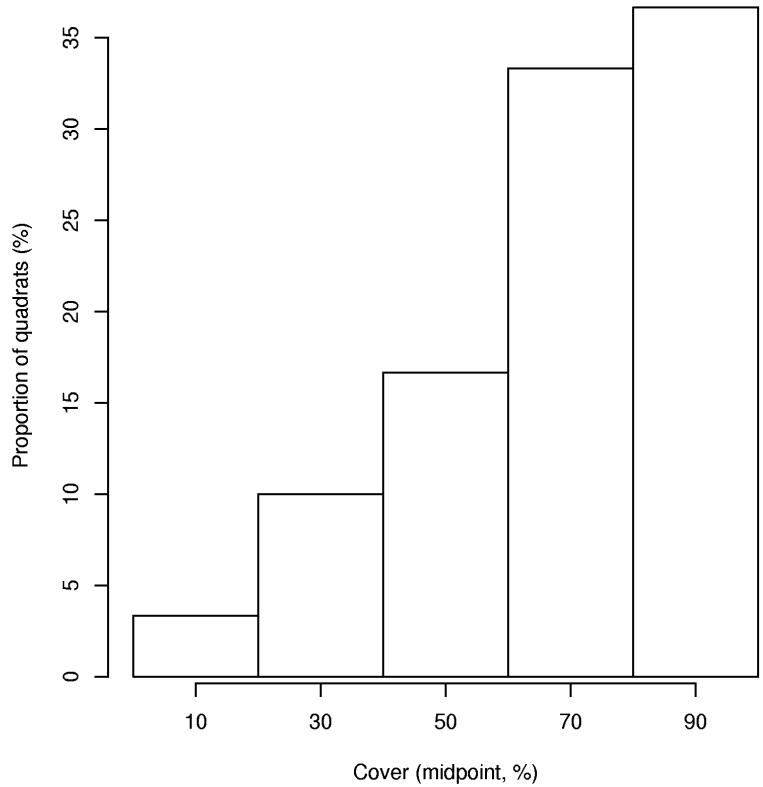
(overall estimates)

Overall Coverage (TFC)	Overall Coverage (CTC)
26.3%	73.9%
(95% CI = 19.1%, 34.5%)	(95% CI = 64.7%, 81.8%)

TFC vs. CTC

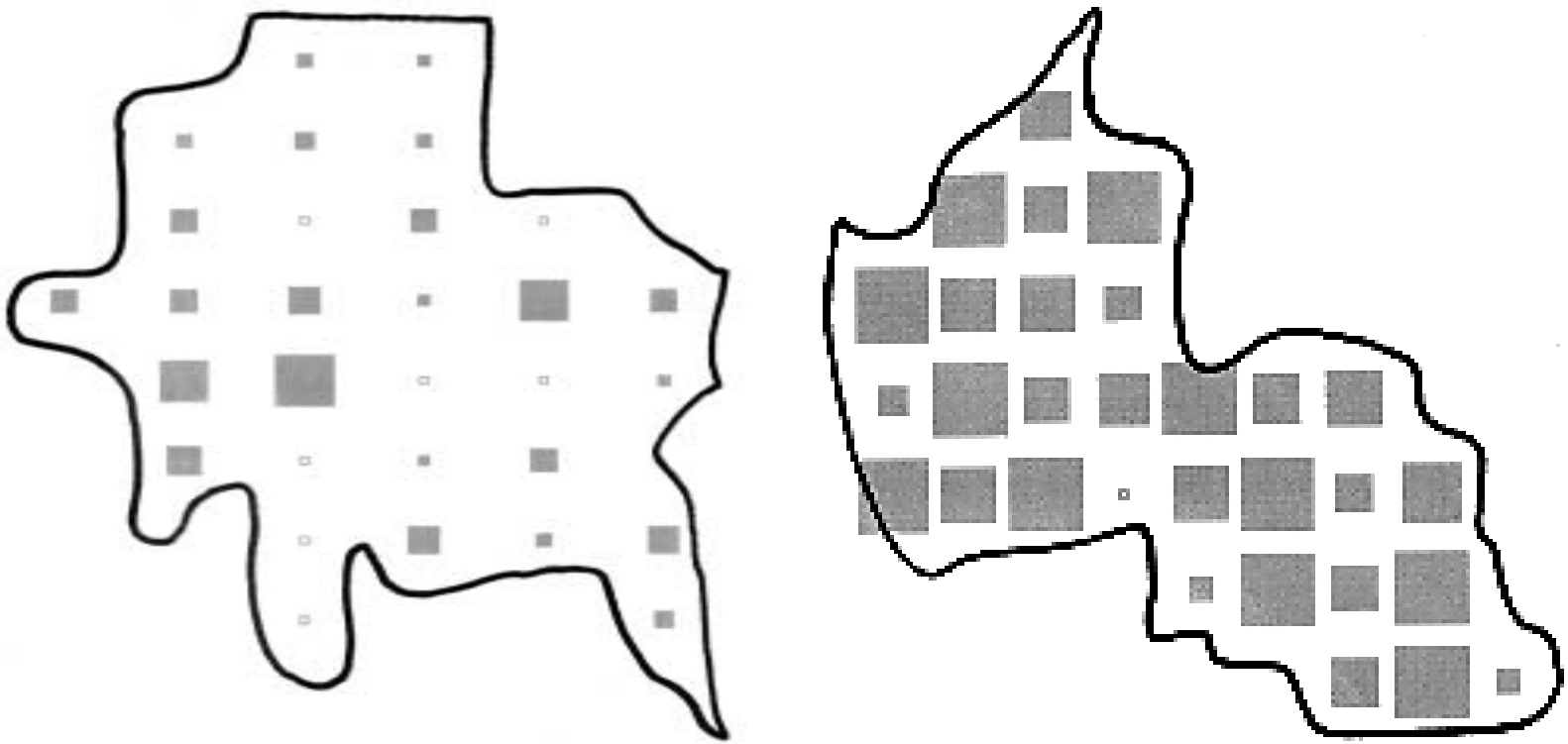
(range of coverage)





TFC vs. CTC

(map of coverage)



Conclusion

- Coverage can be estimated at both small and wide area level.
- Under *study* conditions … CTC provides better coverage than TFC.
- Similar coverage estimates and pattern of coverage achieved by CTC under *program* conditions in Ethiopia.