PSI RESEARCH TOOLKIT

BUILDING RESEARCH CAPACITY

Equity Analysis: Computing the Concentration Index

PSI's Core Values

Bottom Line Health Impact * Private Sector Speed and Efficiency * Decentralization, Innovation, and Entrepreneurship * Long-term Commitment to the People We Serve

Research Division
Population Services International
1120 Nineteenth Street, NW, Suite 600
Washington, DC 20036

Equity Analysis: Computing the Concentration Index

PSI Research Division
2007

Contact Information

Peter Buyungo[1] and Hongmei Yang, PhD[2]

1. Research Director, PSI/Uganda
2. Research Officer, PSI/Washington

For more information, please contact:

Hongmei Yang, PhD
Population Services International
1120 19th Street, NW, Suite 600
Washington, DC 20036

Telephone +1 202 785 0072
Email hyang@psi.org

EQUITY ANALYSIS: COMPUTING THE CONCENTRATION INDEX

LEARNING OBJECTIVES

By the end of this chapter, the reader will be able to:

1. Explain equity and the concentration index.
2. Calculate a concentration index.
3. Compare the difference between two concentration measures.

BACKGROUND

PSI's mission is to improve the health of the poor through the promotion of subsidized products and services. Since ill health is more likely to occur among the poor, the organization's mission is to reduce this occurrence, thereby reducing related inequality. PSI defines *equity* as the absence of a rate and/or frequency difference in the use of a health product or service (given need) by socio–economic status. *Socio–economic status* (SES) can be defined through income, expenditure or asset ownership, education, gender, and residence. When defined as an impact, equity is the reduction or elimination of a rate or frequency difference in the use of health product or service given need. Equity is one of the key measures of social marketing performance. Measuring equity levels will answer various questions for managerial and strategic decision making as shown in Table 1.

TABLE 1: MANAGERIAL AND STRATEGIC QUESTIONS

Managerial Questions	Strategic Questions
• What are the population characteristics of social marketing users? • How does that compare to users of other health services? • Is the marketing plan effective at reducing socio-economic differentials in the practice of risk reducing behavior among those with equal need?	• Does social marketing enhance equity in health service use and other risk reducing behaviors?

How is inequality identified between the poor and the rich in unhealthy behaviors? This question may be answered by examining the distribution of unhealthy behaviors among the different socio-economic levels of a population.

The purpose of this toolkit chapter is to introduce and describe a tool for measuring equity, the *concentration index*. The concentration index provides a means of quantifying the degree of income

related inequality in a specific health variable. The concentration index is defined with reference to the concentration curve, which graphs on the x-axis the cumulative percentage of the sample ranked by living standards beginning with the poorest and on the y-axis the cumulative percentage of the health variable corresponding to each cumulative percentage of the distribution of the living standard variable. The concentration index is defined as twice the area between the concentration curve and the line of equality. In the case, where there is no income related inequality, the concentration index is zero. The convention is that the index takes a negative value when the curve lies above the line of equality (indicating a disproportionate concentration of the health variable among the poor) and a positive value when it lies below the line of equality. *Quantitative Techniques for Health Equity Analysis – Technical Note #7* states that if a health variable is "bad" (e.g., ill health), a negative value of the concentration index means ill health is higher among the poor.

The concentration index can be used to determine if programs are reaching the poor, determine differences in equity among countries, demonstrate how PSI's interventions are changing in their reach of the poor, and determine behaviors with high degrees of inequality among other uses.

HOW-TO-STEPS

1. **Conduct a TRaC survey and include performance review and SES questions.**

   To compute the concentration index, one should have data on the behavior of interest and the socio-economic status of the population. Therefore, PSI programs need to conduct TRaC studies and include questions for the performance review indicators and SES.

   SES can be measured using three approaches: (1) by examining assets, possessions and amenities; (2) by examining income levels; and (3) by examining expenditure levels. In TRaC, the main method of measuring a population's SES is by examining their ownership of assets, possessions, and amenities. It is very important to invest time in the creation of an appropriate list of assets, possessions, and amenities that will make a distinction in the population in question. The best items to be included in the questionnaire should be owned by over 20% of the population but less than 80% of the population.

   In cases where respondents are asked about their levels of income and expenditures as a measure of SES, both actual figures and categories of income are acceptable for the analysis. If a categorical variable is used, make sure there are at least five categories.

2. **Compute SES as a continuous variable.**

   If SES has been measured using questions on income or expenditure, they should already be in continuous form and will not need conversion. If SES is based on assets and possessions, this should be converted into an index using the principle component analysis (PCA) method. However, instead of creating SES categories from the factor analysis, the first factor provides each individual's rank in the socio-economic status distribution. If SES has been measured as a categorical variable with different levels of income/expenditure, make sure that the higher income/expenditure level has a greater value.

3. Compute non-behavior.

The goal is to quantify the degree of income related inequality in negative behaviors; therefore, all performance review indicators that are not already negatively worded should be recoded to provide negative outputs. These include indicators on abstinence, use of condoms, ownership of nets, ownership of ITNs, modern family planning, and treatment of drinking water. All positively coded behaviors need to be transformed into non-behavior. This means users=0 and non-users=1.

> The health variable is computed as bad (i.e., non-behavior or unhealthy behavior) because the goal is to find out whether unhealthy behavior is higher among the poor. Thus, the health variable is coded as bad such as 'does not own a mosquito net'.
>
> By converting all the behaviors into the negative, questions such as "Is there inequity in the unhealthy behavior?" and "Are the poor more likely to experience unhealthy behavior?" are answered.

4. Compute the concentration index and standard error using the syntax below.

This syntax is prepared from the formulas on computing the concentration index for micro data and computing a standard error for the concentration index in *Quantitative Techniques for Health Equity Analysis, Technical Note #7*. Where data are weighted, a weighted fractional rank and covariance needs to be generated.

The parts of the syntax which are **bold** should be customized by country data. Where the syntax below is split into parts with notes, please run it as one for each risk group and behavior combination. In the case of one round data, parts one thru four in the syntax are enough since it will give the concentration index only. In the case of two rounds of data, parts five thru seven are necessary for deriving the standard error, which is required in testing for significant changes in the concentration index over time.

Variables

nonuse = did not use condoms at last sex with a non – marital partner

SES = socio economic status (in continuous format)

Age = respondents age

Sex = respondents sex

Important Note: Please create a folder on one of the drives named 'temp' where the temporary file created using this syntax will be stored. The folder is required for this analysis. This folder has to be a root folder (e.g., u:\temp or h:\temp or c:\temp). Ensure that this part of the syntax is changed in order to make it compatible to the root folder (i.e., u, h, c, p, or etc.) Should problems be encountered in creating this folder, please show this part of the syntax to an IT person who should be able to help create the folder. In the syntax below, the 'temp' folder was created on the 'u' drive (indicated in bold).

TABLE 2: SYNTAX FOR COMPUTING THE CONCENTRATION INDEX AND STANDARD ERROR FOR MICRO DATA

****Part 1. First select the relevant risk group and behavior** 15-24years FEMALE **Condom use at last sex**.**

```
USE ALL.
COMPUTE filter_$= (age>=15 & age <= 24 & sex=1 & (nonuse=0 | nonuse=1)).
VARIABLE LABEL filter_$ ' age >= 15 & age <= 24 & sex=1 & (nonuse=0 | nonuse=1) (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

RANK VARIABLES = SES (A) /RFRACTION into rnkcon /PRINT=YES /TIES=MEAN.
```

**** Part 2. Calculate the: (total number of cases) = n, (mean of dependent variable) = avgnonuse, (mean of rank of SES) = avgrnk.**

```
sort cases by nonuse.
compute nobreak=1 .

aggregate /outfile= ' u:\temp\temp.sav' /presorted /break= nobreak /n=N /avgdv= mean(nonuse)
/avgrnk=mean(rnkcon).

match files file=* /table=' u:\temp\temp.sav' /by= nobreak .
```

**** Part 3. Calculate: covariance, which is cov(yi, Ri), where Ri is the ith individual's fractional rank in the socioeconomic distribution.**

```
compute covar_i=(rnkcon-avgrnk) * (nonuse -avgdv) .

sort cases by covar_i .
aggregate /outfile=' u:\temp\temp.sav' /presorted /break= nobreak /sumcovar= sum(covar_i) .

match files file=* /table=' u:\temp\temp.sav' /by= nobreak .

compute covar=sumcovar/n .
```

**** Part 4. Get the concentration index. **Note: incase you have one round of data, then you only need the concentration index output – using the syntax below.**

```
compute ci=2 * covar/ avgdv .
descriptives var= ci.
```

**** Part 5. Calculate: cumy (cumulative of yi), lagy . But first need to be sure all non-targeted respondents should not be included in the calculation of cumy and lagy.**

```
compute newdv= nonuse.
IF missing(rnkcon) newdv = 99 .
recode newdv (99=sysmis) (else=copy) .

sort cases by rnkcon (A) .
compute indy= newdv / (n * avgdv) .
create cumy= csum (indy) .
COMPUTE lagy = LAG(cumy) .
```

**** Part 6. Calculate: ai and sum of ai-square.**

```
compute a= (newdv/ avgdv) * (2 * rnkcon -1 - ci) + 2 - lagy - cumy .
compute asq=a*a .

sort cases by asq .
aggregate outfile=' u:\temp\temp.sav' /presorted /break= nobreak /sumasq= sum(asq) .

match files file=* /table='u:\temp\temp.sav' /by= nobreak .

** Part 7. Calculate variance of ci and standard deviation of ci .

compute var_ci=sumasq/n**2 - (1+ci)**2/n .
compute std_ci=(var_ci)**(1/2) .

descriptives var=n ci var_ci std_ci .
execute .
```

****Important Note.**
****Delete variables** rnkcon nobreak n avgdv avgrnk covar_i sumcovar covar ci newdv indy cumy lagy a asq sumasq var_ci std_ci **from the dataset. Failure to delete these variables will generate errors of replication of existing variables.**

**Then run analysis for the second risk group.

Outputs

The following example shows the distribution of ill-health/ behavior (i.e., non-use of condoms at last sex with a non-marital partner) according to socio-economic status among female sex workers (FSW) in Nepal.

The first sample output (Table 3) will be produced with one point data and runs parts one thru four of the syntax. The row labeled 'ci' shows the concentration index which is -.2462. Since the result is a negative, this means there is a disproportionate concentration of non-use of condoms among the poor – non-use is higher among the poor FSW.

TABLE 3: SAMPLE OUTPUT– ONE POINT DATA

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
ci	656	-.25	-.25	-.2462	.00000
Valid N (listwise)	656				

When there is more than one round of data, all the parts of the syntax one thru seven will be run. Table 4 shows the results produced. The important information in this output is found in the mean column for rows 'n', 'ci' and 'std_ci'. These three results should be recorded for each time period and used in testing for significant change in the concentration index over time. In this example, n (sample size) = 650; ci (concentration index) = - 0.2462, and std_ci (standard error) = 0.0776.

TABLE 4: SAMPLE OUTPUT– MORE THAN ONE ROUND OF DATA

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
n	650	650	650	650.00	.000
ci	650	-.25	-.25	-.2462	.00000
var_ci	650	.01	.01	.0060	.00000
std_ci	650	.08	.08	.0776	.00000
Valid N (listwise)	650				

Important Note: Whenever the concentration index is a negative, this means the behavior variable being measured is higher among the poor but lower among the rich. A positive concentration index means the variable is higher among the rich but lower among the poor.

5. Use ANOVA to check whether the concentration index has changed significantly overtime.

For two rounds of data, produce the concentration index and the standard error using parts one thru seven the syntax above for each time period. Then use ANOVA to test for significant differences in the concentration index between the two time periods. Use the syntax below to complete the analysis.

Example: For time 1 and 2, input the sample size, concentration index, and standard error generated from the previous analysis into the corresponding parts of the syntax below. In the example below, time 1 sample size (n) = 650, concentration index (mean) = - 0.2462, and standard error (stddev) = 0.0776. In time 2, sample size (n) = 1096, concentration index (mean) = - 0.2951, and standard error (stddev) = 0.1040.

Syntax

```
matrix data variables=time rowtype_  score / factors= time .
begin data
1 n 650
1 mean -0.2462
1 stddev 0.0776
2 n 1096
2 mean -0.2951
2 stddev 0.1040
end data .

oneway score by time /matrix= in(*) .
```

In the SPSS output (Table 5), examine the ANOVA table. In this table, check whether the result between groups is significant. This table below shows significance of p<.001. This is the level of significance reported for differentiating between the concentration indices in the two time periods. Since the concentration index at time two is bigger in its value, it seems that inequity is getting worse compared to time one.

TABLE 5: SPSS OUTPUT TABLES

Descriptives

score

	N	Mean
1	650	-.246200
2	1096	-.295100
Total	1746	-.276896

ANOVA

score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.976	1	.976	108.023	.000
Within Groups	15.752	1744	.009		
Total	16.727	1745			

CASE EXAMPLES AND LESSONS LEARNED

At this time there are not case examples or lessons learned to be documented.

QUALITY IMPROVEMENT CHECKLIST

FIGURE 1: CHECKLIST FOR CALCULATING EQUITY

- ☐ Does continuous or categorical data on SES exist? If there is categorical SES data, do higher values indicate higher social-economic status?
- ☐ Have all the health variables been recoded to reflect ill health?
- ☐ Has a folder called ‘temp’ been created on one of the drives?
- ☐ Have the names of variables been changed in the syntax to reflect country specific data? (Shown in **bold** in the syntax box in this document.)

APPENDIX

Recommended Reading

World Bank (n.d.) *Quantitative techniques for health equity analysis—technical note #7: the concentration index.* Retrieved November 2, 2007 from http://www1.worldbank.org/prem/poverty/health/wbact/health_eq_tn07.pdf