Development of competition indicators in the Norwegian general practice:

Constructing a postal code-specific Herfindahl-Hirschman index applying STATA software

Xing Chen
Geir Godager
Institute of health and society,
Department of Health Management
and Health Economics,
University of Oslo

**UNIVERSITY
OF OSLO**
HEALTH ECONOMICS
RESEARCH PROGRAMME

Working paper 2011: 1

HERO

Development of competition indicators in the Norwegian general practice:

Constructing a postal code-specific Herfindahl-Hirschman index applying STATA software

Xing Chen
Institute of health and society,
Department of Health Management and Health Economics, University of Oslo

Geir Godager
Institute of health and society,
Department of Health Management and Health Economics, University of Oslo
P.O. Box 1089 Blindern, NO-0317 Oslo, Norway

E-mail: geir.godager@medisin.uio.no

March 14th 2011

**Health Economics Research Programme at the University of Oslo
HERO 2011**

ABSTRACT

The purpose of this text is to describe a procedure that can be used to construct a Herfindahl-Hirschman index of market concentration based on the following two pieces of information:

a) An output measure of individual producers located in areas identified with a postal code.

b) A matrix containing information about the distances between the (centers of) the areas defined by postal codes.

The examples presented in this text are from the market for general practitioners' services in Norway, but applying the suggested procedures to other output measures should be straight forward. The STATA commands necessary to construct the index are provided.

1 Background and motivation

The Herfindahl-Hirschman index is a measure of market concentration named after Orris Herfindahl, who applied the index in is his doctoral dissertation (Herfindahl, 1950), and Albert Hirschman, who presented the index in a book (Hirschman, 1945)[1]. In the literature the index is most often expressed in the form of the sum of the squared market shares of individual firms in the market[2]: $\sum s_i^2$, ranging from zero to one if the market shares are expressed in terms of proportions and from zero to ten thousand if the market shares are expressed in terms of percentages. The index is a widely used measure of market concentration, also in markets for health care services (Pauly, 2004). Application of this index in the context of the market for general practitioners' services is not very common in the literature, perhaps due to difficulties in measuring the output. General practice in Norway is organized as regular GP scheme with patient lists. The number of patients listed in each practice may thus be used as an output measure. This output measure is available from the Norwegian primary physician database. A part of the project "Development of competition indicators in the Norwegian general practice" is to construct the Herfindahl-Hirschman index for areas defined by five and ten kilometer radius from the center of areas identified with a postal code, by applying this output measure. Not much is known about the level of competition in Norwegian general practice, even though the competitiveness of this market is likely to have important influence on aspects of service delivery such as service quality and the number of referrals to secondary care. Development of this commonly used measure of competition may hopefully contribute to future research aiming to investigate the impact of competition on outcome measures in Norwegian general practice.

---

[1] For a discussion on the origin of the index, see Hirschman (1964).
[2] In Hirschmans version, the index was given by the square root of this sum, see Szpiro (1987) for a discussion on the properties of the two versions of the indexes.

2 Constructing the Herfindahl-Hirschman index in four steps

We assume there are *J* unique postal codes. In the following examples below we have J equal to 10. For these 10 areas, there exists a distance matrix indicating the travel distances in kilometers between centers of the postal code areas. An illustrative example of the matrix is given in Table 1. We see that the distance between postal codes 15 and 20 is 1.2 kilometers while the distance between postal codes 17 and 20 is 10.9 kilometers. We provide an example to illustrate how defining market sizes in terms of kilometers determines which GPs are considered to be part of the same market, and which GP is considered to be part of different markets: If the market size is defined by a radius of within 10 kilometers from a certain postal code, GPs located in postal code 15 will be considered to be part of the same market as GPs in postal code 20, whereas GPs in postal code area 17 will not be considered to be part of the same market as GPs in postal code 20.

Table 1. Distance matrix in wide format

	Postal code									
frompostalcode	**10**	**15**	**16**	**17**	**20**	**21**	**22**	**23**	**24**	**25**
10	0	1.9	5.3	9.5	1.7	1.7	1.7	10	10	1.5
15	1.9	0	6.8	11	1.2	1.2	1.2	9.3	9.3	.9
16	5.3	6.8	0	5.9	6.7	6.7	6.7	14.2	14.2	6.5
17	9.5	11	5.9	0	10.9	10.9	10.9	17.4	17.4	10.8
20	1.7	1.2	6.7	10.9	0	0	0	8.6	8.6	.2
21	1.7	1.2	6.7	10.9	0	0	0	8.6	8.6	.2
22	1.7	1.2	6.7	10.9	0	0	0	8.6	8.6	.2
23	10	9.3	14.2	17.4	8.6	8.6	8.6	0	0	8.8
24	10	9.3	14.2	17.4	8.6	8.6	8.6	0	0	8.8
25	1.5	.9	6.5	10.8	.2	.2	.2	8.8	8.8	0

There are several producers located at the center of postal code j. We refer to total output in postal code j as N_j, while the output of producer i in postal code j is denoted by n_{ij}. In this application the output of an individual general practitioner (GP) is the list size of his patients. A single postal code may be referred to a as a market itself, and we express the market share of provider i in postal code j as n_{ij}/N_j. There are I_j Producers in postal code j. The Herfindahl, when considering the single postal code area as being the market can be expressed as[3]:

$$\sum_i^{I_j} \left(\frac{n_{ij}}{N_j} \right)^2$$

We assume every market is shaped as a circle whose radius is 5 km. When a market is defined to be postal code A plus postal codes closer than 5 km, the market identified by postal code A will include GPs in postal code B also. An important element to be carefully addressed is the fact that GPs will frequently be included in several such circles at the same time. In a market consisting jointly of postal code A (with two providers 1 and 2) and B (with two providers 3 and 4), the Herfindahl index for this area may, by definition, be expressed as.

$$1) \quad H_{AB} = \left(\frac{1}{N_A + N_B} \right)^2 \left(n_{1A}^2 + n_{2A}^2 + n_{3B}^2 + n_{4B}^2 \right)$$

When the market includes GPs in several postal-code areas, we can construct the Herfindahl for the joint area by applying a step-wise procedure providing the information needed to compute 1).

The first step: **Constructing variables that describe the J individual postal code areas such as N_A and the sum $n_{1A}^2 + n_{2A}^2$**

- Sum the output of individual providers within each postal code   (we get in particular : N_A)

---

[3] In this note we apply throughout the normalization where the index takes values between zero and one.

STATA command ->by postalcode, sort: egen Nj= total(listsizeGPi)

Square the output of each provider within the postal code and sum (we get in particular $(n_{1A}^2 + n_{2A}^2)$

STATA command   -> gen   listsizeGPi _sq= listsizeGPi ^2

->by postalcode, sort: egen sumlistGPisq_j = total(listsizeGPi _sq)

- Collapse the file and save it:

STATA command->collapse sumlistGPisq_j   Nj , by (postalcode)

->save c:\collapsedfile.dta

We now have a file describing individual postal code. It may look something like Table 2:

Table 2

postalcode	Nj	sumlistGPi sq_j
10	7500	9375000
15	3500	4083333
16	8000	9142857
17	4000	5333333
20	3000	3000000
21	1500	1125000
22	1500	2250000
23	9500	11281250

The second step: Preparing the drive distance matrix so that relevant postal codes are linked.

Arrange the distance matrix so it is stored in long format by applying the STATA command reshape.   An extract may thus look like Table 3 for the first lines describing all the postal codes related to postal code 10. If the aim is to construct a Herfindahl index based on a radius of x km, make sure that combinations of postal codes that results in travel distances> x kilometers are deleted.

Table 3: Distance matrix only containing distances<x.

frompostalcode	postalcode	distance
10	10	0
10	15	1.9
10	20	1.7
10	21	1.7
10	22	1.7

The third step:

- Starting with the distance matrix in long format (illustrated by table 3) merge the file containing Total output (Nj) and the sum of the squared output for individual providers in each individual postal code area (*sumlistGPi _sqj*) (illustrated by table 2). To the distance matrix by using the variable *postalcode* as key.[4]

  STATA command-> **use c:\longdistancematrix.dta, clear**

  merge m:1 postalcode using collapsedfile.dta

- Perform new computations of output in joint areas such as $N_A + N_B$.[5]

  STATA command ->by frompostalcode, sort: egen Nsum= total(Nj)

  And then square of this sum:

  STATA command -> gen Nsumsq= Nsum^2

- And the relevant sums of squared list sizes:

  STATA command ->by frompostalcode, sort: egen nsqaredsum= total(sumlistGPisq_j)

The final step: Compute the Herfindahl indices

- STATA command ->gen herfindahl= nsqaredsum/ Nsumsq

- Now this file can be collapsed to produce a file with a Herfindahl corresponding to

---

[4] Note that both the files need to be sorted according to the variable *postalcode*.
[5] Note that computations are now performed within the variable *frompostalcode*. Hence, the total output is

each postal code:

3 Application: Brief description of the development of the Herfindahl-Hirschman index in Norwegian general practice.

Figure 1. Development of the Herfindahl-Hirschman index in the period 2001-2010 using observations in July each year. Market defined by 5 kilometer radius. Observations are weighted by size of GP output, hence the median describe the median in the Norwegian population.

---
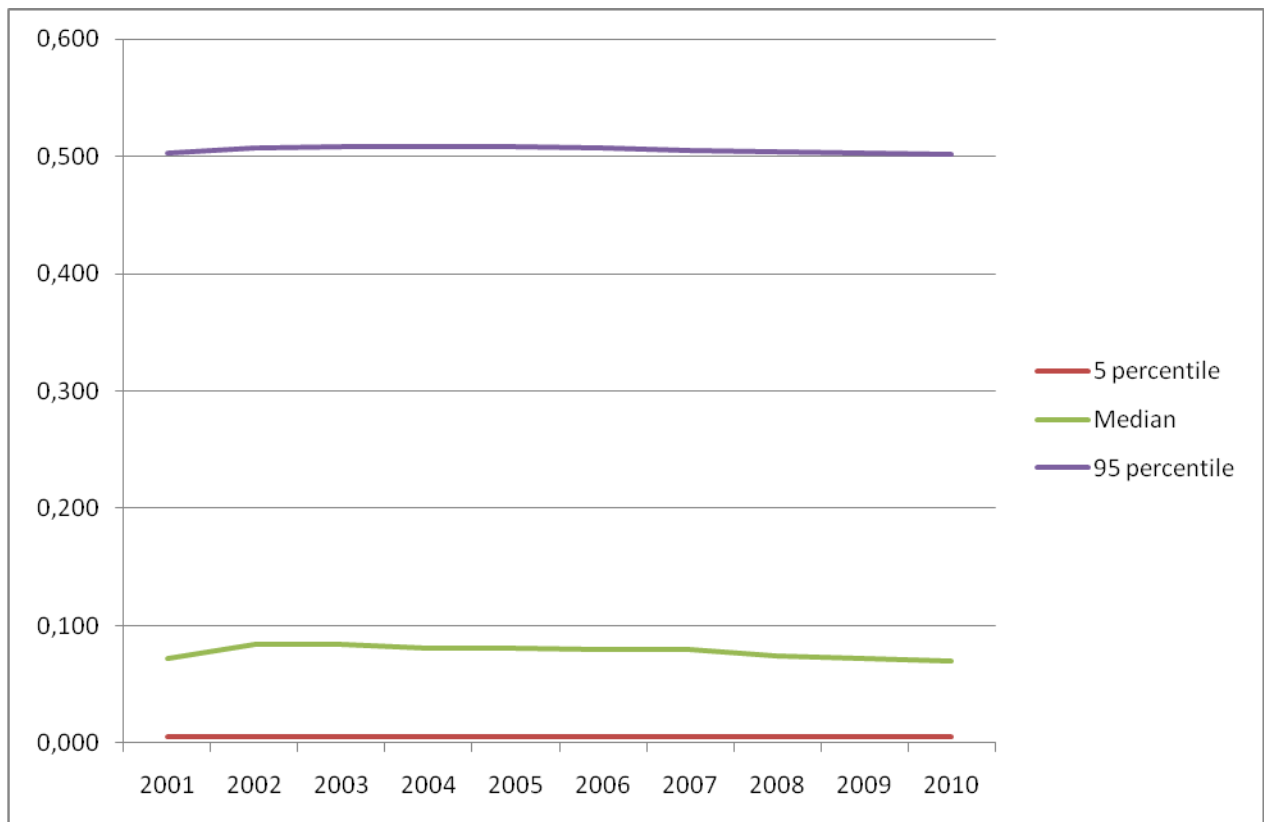
computed within the joint are of zip code j and all neighbor zip codes closer than x kilometers.

Figure 2. Histogram of the Herfindahl-Hirschman index in July 2001. Market defined by 5 kilometer radius. Observations are weighted by size of GP output.
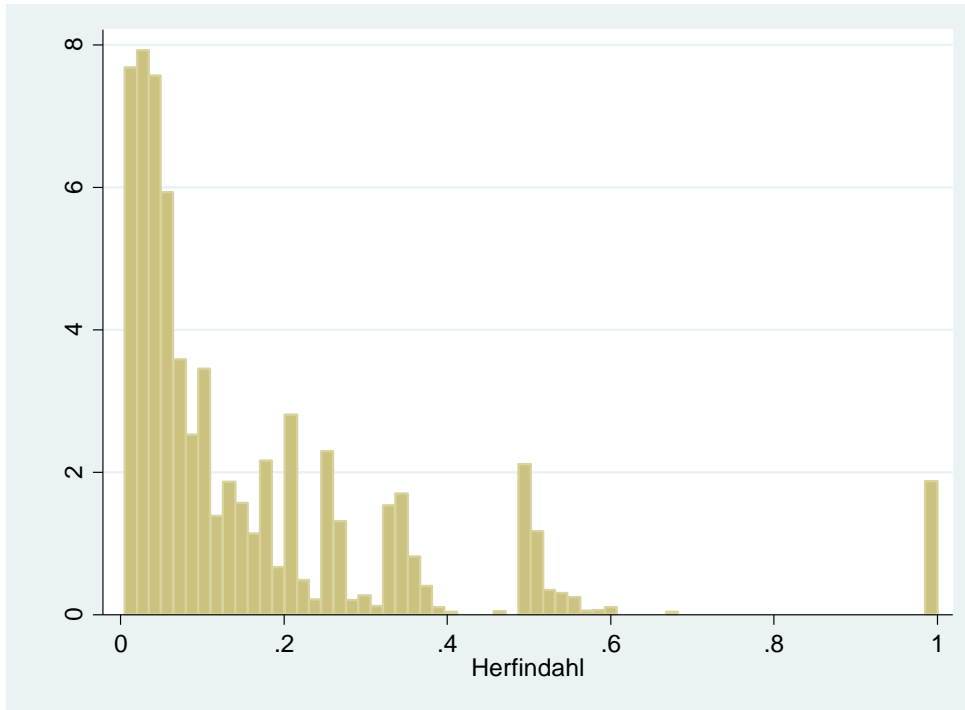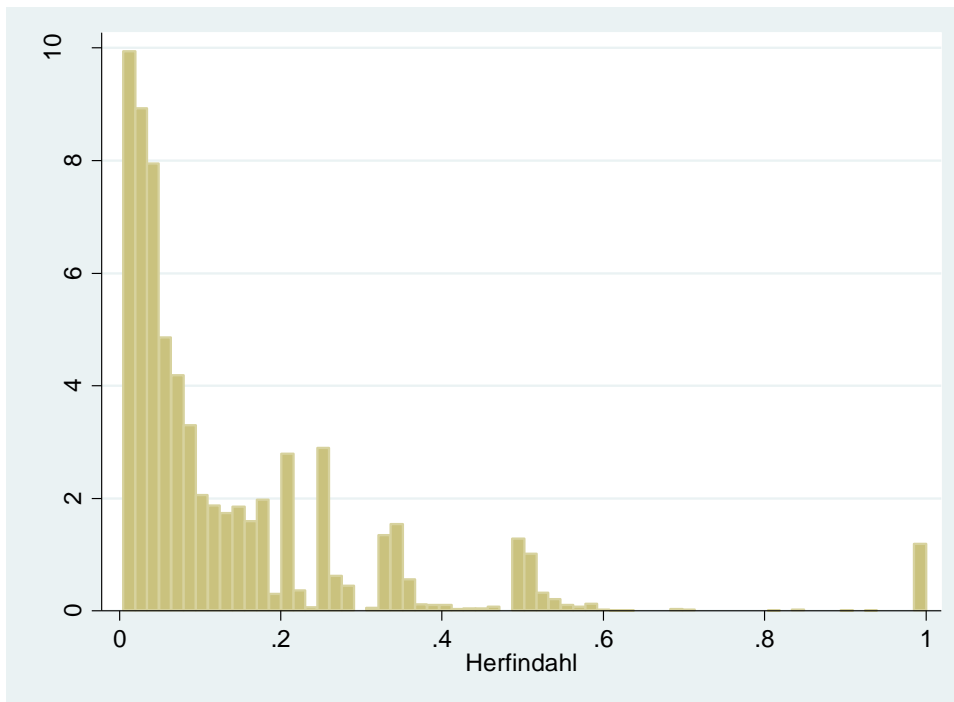

Figure 3. Histogram of the Herfindahl-Hirschman index in July 2010. Market defined by 5 kilometer radius. Observations are weighted by size of GP output.

References

Herfindahl, Orris C., 1950.   Concentration in the U.S. Steel Industry. Unpublished

doctoral dissertation, Columbia University.

Hirschman, Albert O., 1945. National Power and the Structure of Foreign Trade.

Berkeley.

Hirschman, Albert O., 1964. "The Paternity of an Index", *The American Economic Review*, **54**(5), p. 761

Szpiro, George G., 1987: "Hirschman versus Herfindahl: Some topological properties for

the use of concentration indexes", *Mathematical Social Sciences*,**14**(3),

P.299-302

Pauly, Mark V.,2004. "Competition in Medical Services and the Quality of Care: Concepts

and History", *International Journal of Health Care Finance and Economics*, 4,

113–130