

Fractal analysis of time series in epidemiology: Is there information hidden in the noise?

Jan F. Nygård^{1,2} and Eystein Glatre^{1,3}

1) Cancer Registry of Norway, Oslo, Norway

2) University of Oslo, Institute of Behavioural Science in Medicine, Oslo, Norway

3) Norwegian School of Veterinary Science, Oslo, Norway

Correspondence: Jan F. Nygård, Kreftregisteret, Montebello, N-0310 Oslo, Norway

Telephone: +47 22 45 13 02 Telefax: +47 22 45 13 70 E-mail: jan.nygard@krefregisteret.no

ABSTRACT

Monthly total cancer incidence rates for women and monthly thyroid cancer incidence rates for women were retrieved from the Cancer Registry of Norway for the period 1953-2000, that is, 246-1061 cases per month and a total of 331 196 cases, and 0-24 female thyroid cases per month, a total of 5045, during this half-century. De-trended fluctuation analysis *ad modum* Peng et al., briefly explained in the article, are used to show that these two morbidity curves are statistical fractals, self-affine and non-Euclidean. The self-affinity parameter $\alpha = 0.87$ of the female thyroid cancer curve indicates persistent long-range power-law time-correlations of the curve fluctuations. This finding indicates that exposure has occurred in the female population leading to the occurrence of thyroid cancer cases spread over a wide period. For the total female cancer curve, on the other hand, the self-affinity parameter $\alpha = 1.35$ indicates time-correlations, but not the long-range, power-law type.

NORSK SAMMENDRAG

Månedlig kreftinsidensrate for kvinner og månedlig skjoldkjertel kreftinsidensrate for kvinner i perioden 1953-2000 fra Kreftregisteret er benyttet i denne artikkelen. For alle kvinner varierte antall tilfeller mellom 246 og 1061 per måned og totalt 331 196 tilfeller, for skjoldkjertel varierte antall tilfeller fra 0 til 24, og 5034 tilfeller totalt. Vi benyttet "de-trended fluctuation" analyse, kort beskrevet i artikkelen, for å vise at begge insidenskurvene er statistiske fraktaler, egen-affine og ikke-Euklidske. Den egen-affine parameteren $\alpha = 0,87$ for skjoldkjertelkreft viser vedvarende langt-rekkende potenslov tids-korrelasjoner for insidensraten. Dette indikerer at eksponering for risikofaktorer i den kvinnelige befolkningen er relatert til insidensrater over et vidt tidsrom. Totalinsidenskurven for kvinner med egen-affinitets parameter $\alpha = 1,35$ viser korrelasjon over tid, men ikke av den langt-rekkende potenslov typen.

INTRODUCTION

Randomness is the label assigned to the enigmatic and relatively unknown universe of observation-disturbing influence that all scientists experience always and everywhere. However, research has revealed that there exist at least some specific classes of randomness within this universe, with relevance for biology and medicine (1-3). In spite of that, randomness, materialised as noise, has attracted little attention among epidemiologists. We still tend to regard measurement variation and all other types of noise as something one should preferably get rid of in order to detect the underlying 'truth'. Thus, the fluctuations one can discern and describe as part of secular morbidity and mortality curves are treated as an annoying feature; by most epidemiologists, noise is regarded as caused by anonymous, completely uninteresting, naturally occurring 'error-producing' processes, which should be eliminated. The epidemiologist's aim has therefore been to disclose central tendencies in the series of rates and

estimate measures of deviation from this tendency. Prediction of future incidence rates, based on extending observed tendencies into the future, has also been a highly appreciated research activity.

Fractal statistics represents something quite different. This chapter of statistics gives the scientist extended opportunities to investigate aspects of nature that have hitherto appeared too complex and inaccessible. Fractal methods provide insights into the mechanisms of pattern formation in phenomena such as bacterial growth patterns, fracture formation, brain waves, heartbeat dynamics, and fluctuating morbidity curves.

With regard to specific classes of randomness, a sub-class is the time-correlated noise (4). For this sub-class, noise at a given time is associated to noise at a different time, e.g. pink noise or 1/f noise indicates long-range correlations, implying complex, non-linear processes that generate fluctuations on a wide range of time scales.

Fractal methods also open up for the study of self-organised criticality (SOC), a possibly widespread

phenomenon where nature is perpetually out of balance, but organised in a poised (critical) state in which anything can happen. To disclose SOC within the realm of epidemiological research is a great challenge. The fact that properties of noise may depend on the phenomenon being studied, makes it relevant to ask whether noise itself might be the most prominent feature of many phenomena (5).

It is the purpose of this study to present fractal methods for the study of time-series and their fluctuations and apply these methods on secular series of cancer incidence rates.

MATERIAL AND METHODS

Cancer incidence rates on a monthly basis were retrieved from the Cancer Registry of Norway for the period 1953 to 2000. 1) The series of incidence rates 1953-2000 for all cancers among Norwegian females and 2) the series of incidence rates 1953-2000 for thyroid cancer among Norwegian females were both subjected to fractal investigation and supplemented with a simple linear regression analysis for trend.

Fractal objects

Introduction to fractals may be found in several textbooks (6-8). Briefly, for objects to be fractal they must satisfy two criteria, that of being self-similar or self-affine and that of having a fractional, that is non-Euclidean, dimension. Self-similarity/-affinity means that the object is composed of subunits and sub-subunits on multiple levels that exactly or statistically resemble the structure of the whole object (8). Isotropic fractals are self-similar, i.e. they are invariant under isotropic scale transformations. In contrast, objects that are invariant under anisotropic scale transformations belong to the class of self-affine fractals. We call an exactly self-similar/-affine fractal structure deterministic; on the other hand, fractal structures that are statistically self-similar/-affine are often called statistical fractals. In epidemiology, we are working mainly with statistical fractals.

A solid cube, for instance, is self-similar since it can be divided into subunits of eight smaller solid cubes that resemble the large cube, and so on. However, the cube is not a fractal object because it is three-dimensional; its Euclidean (integer) dimension is 3. On the other hand, so-called Cantor 'dust' sets are self-similar from level to level and their dimensionality is fractional. The triadic variant, for instance, has the non-Euclidean dimension 0.631.

Self-affinity parameter

In principle, one can test whether a two-dimensional curve is statistically self-affine by taking a window of the curve and rescale it along the abscissa and ordinate axes to the size of the original curve, then comparing the statistical properties of the rescaled curve with

those of the original. The statistical properties are equal in case of self-affinity. One uses in practice, however, a weaker criterion – that it is sufficient for the means and variances of the two curves to be the same. In mathematical terms the original curve is self-affine if

$$y(t) \equiv a^{\square} y\left[\frac{t}{a}\right]$$

implying that first and second order moments on both sides of the equation are identical. If so, the exponent \square is called the self-affinity parameter; cf. the Lipschitz-Hoelder rescaling exponent (8).

Mapping time series

Self-affine time series with $\square > 0$ will be unbounded, that is, fluctuations on large observation windows are significantly larger than those of smaller windows. In contrast, most time series produced by the life sciences are bounded, that is, they cannot have fluctuations exceeding certain values irrespective of the length of the series. This represents an obstacle to their fractal analysis that can only be circumvented by instead investigating the accumulated time series of the original series. This so-called mapping is a crucial step in fractal time series analysis (9,10).

De-trended fluctuation analysis

De-trended fluctuation analysis (DFA) is the method of choice for extracting the self-affinity parameter, \square , if it exists, from given epidemiological time series. DFA is a modified root mean square analysis of random walk with advantages over conventional methods such as for instance spectral analysis and Hurst analysis. This superiority of DFA is first and foremost due to the fact that DFA can be applied to bounded and to non-stationary time series, especially slowly varying trends. DFA permits the detection of self-affinity embedded in non-stationary time series, and avoids the spurious detection of self-affinity that may be an artefact of extrinsic trends. DFA has been applied successfully to a wide range of simulated and physiological time series in recent years. It is, though, not designed to handle all possible non-stationary real-world series (11).

The first step in DFA is the mapping of the original incidence rate (IR) time series to an aggregated one, that is

$$y(k) = \sum_{i=1}^k [IR(i) \square IR_{ave}]$$

IR(i) is the IR at the i th point in the series and IR_{ave} = average IR of the IR-series. Next, one measures the vertical characteristic scale of the accumulated time series by dividing the latter into boxes of equal length, n . In each box a least-squares line is fit to the data (the box trend). The ordinate of the straight-line segments

is denoted $y_n(k)$. Thereafter, the accumulated series, $y(k)$ for observation k , is de-trended by subtracting the local trend, $y_n(k)$, in each box. The detrending is similar to, for instance, loess-smoothing, and $F(n)$ is the residual standard deviation after smoothing.

For box size, n , the characteristic fluctuation size is calculated by

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2}$$

N is the total length of the series. The same calculation is repeated over different box sizes, n . The self-affinity parameter (α) is defined as the slope of the regression line for all points $(\log(n), \log[F(n)])$, giving the power-law relation $F(n) = kn^{-\alpha}$ (11,12).

Relationship between self-similarity and autocorrelation functions

Introductory texts to time-series include the following books (13,14). The relationship between self-similarity/-affinity and autocorrelation functions is described briefly in table 1 (adapted from (11)). It is outside the scope of this article to describe in detail the autocorrelation.

Non-Euclidean dimension

For fractals in Euclidean dimension, E , the fractal dimension D is given by $D = E + 1/2(3 - \alpha)$. Using $\alpha = 2\beta - 1$, one gets $D - E = 2 - \beta$ in which β is the self-affinity parameter. This formula is valid for $0 < \beta < 2$, thus including noise from white up to brown (6,15,16).

RESULTS

Figure 1 presents the time series of monthly incidence rates for total cancer 1953-2000 among Norwegian women. The general, practically linear, upward trend

of this series, and the fluctuations above and below this straight-lined trend are the most conspicuous features of figure 1.

Linear regression analysis confirms the steep upward trend; it shows, cf. table 2, an increase of 0.52 per 100,000 per year. Figure 2 shows the monthly, thyroid cancer incidence rates for Norwegian women during 1953-2000. Here the fluctuations seem to take place above and below a seemingly wave-like line with a peak in the 1980's.

Table 2 presents the outcome of the fractal analysis and shows that the de-trended time series, both the total female cancer incidence curve and the female thyroid cancer incidence curve, are non-Euclidean with self-affinity parameters estimated to 1.35 and 0.87, respectively. The curves are, in other words, statistical fractals. The self-affinity parameter of the female thyroid cancer curve indicates persistent long-range power-law time-correlations of the noise. For the noise of the total female cancer curve, on the other hand, the self-affinity parameter indicates time-correlations, but not the long-range, power-law type, cf. table 1.

DISCUSSION

In this paper, we have demonstrated that two important time series, the female total cancer and female thyroid cancer morbidity curves from 1953 to 2000, represent non-linear processes. For this period there is a clear, upward, linear trend of the female total cancer morbidity curve; its fluctuations are correlated, but not of the long-range power-law type. With regard to the female thyroid cancer curve, on the other hand, there is no linear trend over 1953-2000, but the curve fluctuations shows interesting long-range power-law correlation.

The increasing cancer morbidity among Norwegian women should probably be evaluated against the fact that during this period several competing causes of death have been declining, allowing the population to

Table 1. The relationship between the self-affinity parameter (α) of an accumulated time series and the autocorrelation function $C(\Delta)$, of the original (non-accumulated) signal.

1. For white noise, where the value at one instant is completely uncorrelated with any previous values, the integrated value $y(k)$, corresponds to a random walk and therefore $\alpha = 0.5$. The autocorrelation function is 0 for any time lag not equal to zero. Many natural phenomena are characterized by short-term correlations with a characteristic time scale, Δ_0 , and an autocorrelation function, $C(\Delta)$, that decays exponentially, i.e. $C(\Delta) \propto \exp(-\Delta/\Delta_0)$. The initial slope of $\log F(n)$ vs. $\log n$ may well be different from 0.5, but α will approach 0.5 for large window sizes.
2. An $0.5 < \alpha \leq 1$ indicates persistent long-range power-law correlations, $C(\Delta) \propto \Delta^{-\alpha}$. The relation between α and β is $\alpha = 2 - 2\beta$. Note also that the power spectrum, $S(f)$ of the original (non-integrated) series is also of a power-law form, i.e., $S(f) \propto 1/f^\beta$ because the power spectrum density is simply the Fourier transform of the autocorrelation function, $\beta = 1 - \alpha = 2\beta - 1$.
3. When $0 < \alpha \leq 0.5$, power-law anti-correlations are present such that large values are more likely to be followed by small values and vice versa.
4. When $\alpha > 1$, correlations exist but cease to be of a power-law form, $\alpha = 1.5$ indicates brown noise, the integration of white noise.

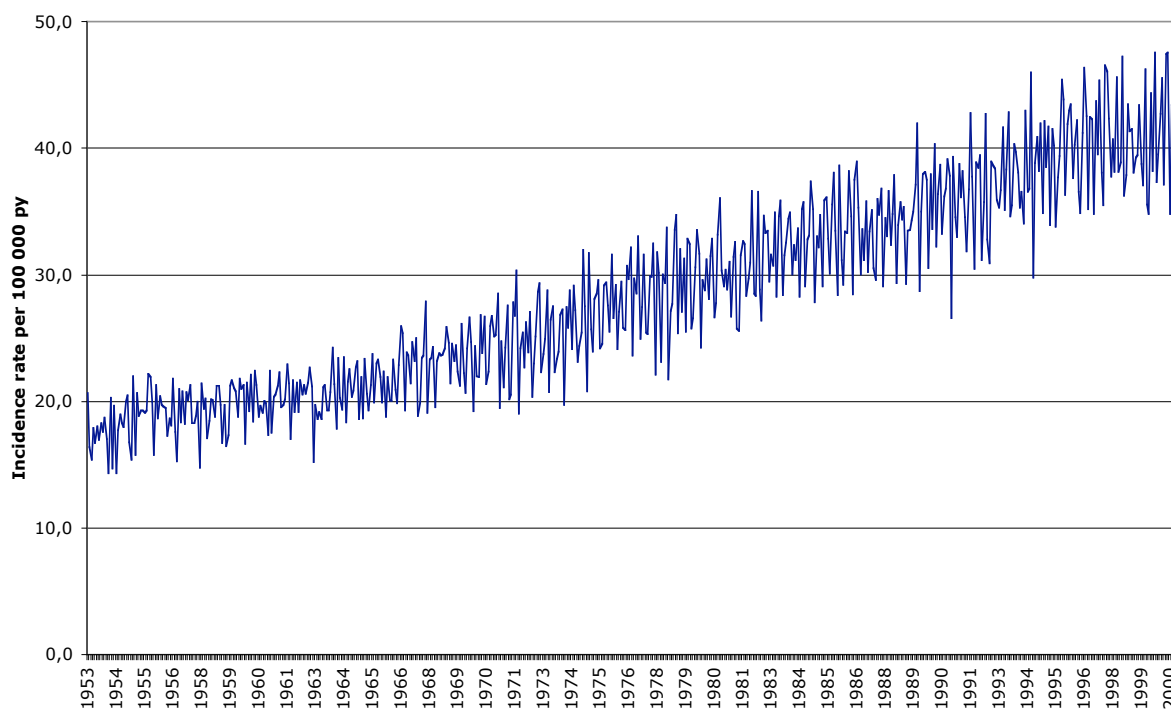


Figure 1. Monthly incidence rates 1953-2000 for total cancer among Norwegian women.

Table 2. Analysis of the incidence rate series 1953-2000 for total cancer and thyroid cancer among women in Norway.

| Cancer type | Self-affinity parameter (\square) | 95% CI of \square | Dimension | OLS regression | 95% CI |
|------------------|---------------------------------------|---------------------|----------------------|----------------|---------------|
| Total, females | 1.35 | 1.31–1.39 | Non-Euclidian (0.65) | 0.52 | 0.50–0.54 |
| Thyroid, females | 0.87 | 0.84–0.90 | Non-Euclidian (1.13) | 0.0078 | 0.0067–0.0089 |

grow older and thereby attract more cancer. However, linear trend is no part of the fractal and will consequently not be discussed further in this paper.

Epidemiological fractals like the cancer morbidity curves in figure 1 & 2 are, in our opinion, the observable representation of underlying complex activities, first and foremost, cancer-producing and diagnosis-producing activities in society. The observed curve fluctuations (figure 1 & 2) are briefly speaking, correlated signals from a highly complex source; not something one should preferably get rid of as noise that obscures a clear view of the central tendency. The long-range power-law noise-correlation disclosed by DFA for the female thyroid cancer curve can be seen as manifestation of a positive exposure-cancer relationship where exposure leads to the occurrence of thyroid cancer cases spread out over a wide period due to individually varying latency periods. Naturally, one cannot expect to disclose a similar feature of the all female cancer curve including all types of cancer in varying proportions over time.

It is frequently the concern of investigators that the length of the time series they want to study by means

of DFA is too short for the method. Both cancer morbidity time series of the approximate length of 575 are well above the lower limit of 50-60 for acceptable DFA analysis. There is no incidence rate of zero in the all female cancer series and one, only, in the female thyroid cancer series.

Physicists' study of surface and interface form and growth has led them to conclude that the rescaling exponent or self-affinity parameter, \square , can be viewed as an indicator of the 'roughness' of the surface and therefore also as a 'roughness-indicator' of time series: the larger the value of \square , the smoother the time series (4). In agreement with this one observes that the total female cancer curve with $\square = 1.35$ is smoother than the female thyroid cancer curve with $\square = 0.87$.

We made an introductory remark about SOC, a phenomenon that we expect to be widespread in biology. This relates to the fact that $1/f$ noise, one of the main characteristics of SOC, can be identified everywhere in nature. It is consequently tempting to guess that SOC is ubiquitous, too. On this background it is interesting to note that the fluctuations of the thyroid cancer incidence curve represent $1/f$ noise or practi-

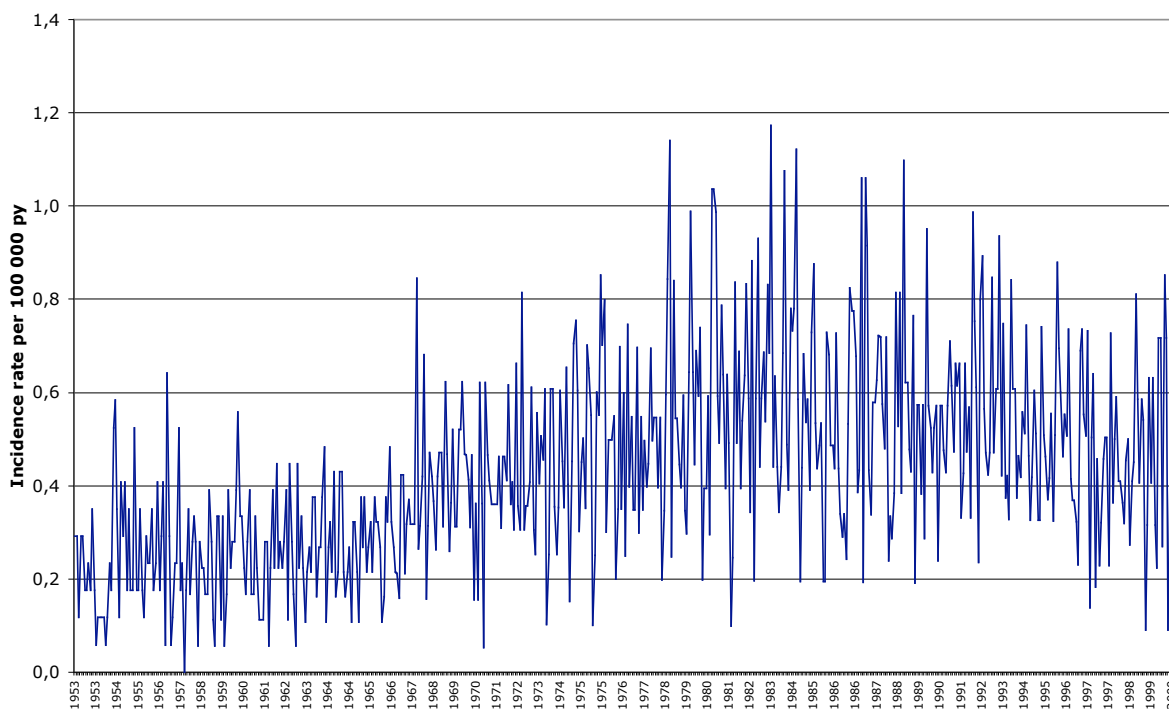


Figure 2. Monthly incidence rates 1953-2000 for thyroid cancer among Norwegian women.

cally so. Incidence rate spikes or clustering would not thus be an anomaly. Clustering in geographical areas or in an occupational environment, could thus not be ascribed to a random event and be dismissed as a freakish incidence, but should rather be viewed as a manifestation of the complexity underpinning the subject under study.

We would like to conclude this little paper by

quoting Manfred Schroeder (7): ‘Self-similarity is, in fact, one of the decisive symmetries that shape our universe and our efforts to comprehend it’.

ACKNOWLEDGEMENT

We thank Aage Johansen for providing us with the monthly cancer incidence rates from the Cancer Registry.

REFERENCES

1. Coffey DS. Self-organization, complexity and chaos: The new biology for medicine. *Nat Med* 1998; **4** (8): 882-5.
2. Stark J, Hardy K. Chaos. Useful at last? *Science* 2003; **301**: 1192-3.
3. Goldberger AL. Fractal variability versus pathologic periodicity: complexity loss and sterotypy in disease. *Perspect Biol Med* 1997; **40** (4): 543-61.
4. Barabási A-L, Stanley HE. *Fractal concepts in surface growth*. Cambridge: Cambridge University Press, 1995.
5. Glatte E, Nygård JF. Fractal meta-analysis and ‘causality’ embedded in complexity: Advanced understanding of disease aetiology. *Nonlinear Dynamics Psychol Life Sci* 2004 (In press).
6. Mandelbrot BB. *The fractal geometry of nature*. San Fransisco: WH Freeman, 1977.
7. Schroeder M. *Fractals, Chaos, Power Laws. Minutes from an infinite paradise*. New York: Freeman and Company, 1991.
8. Feder J. *Fractals*. New York: Plenum Press, 1998.
9. Hurst HE. Long-term storage capacity of reservoirs. *Trans Am Soc Civil Engr* 1953; **166**: 770-99.
10. Kolmogorov AN. The local structure of turbulence in incompressible viscous fluids for very large Reynolds numbers. *Dokl Akad Nauk SSSR* 1961; **30**: 9-13 (Reprinted in *Proc R Soc Lond A* **434**: 913).

11. Peng CK, Hausdorff JM, Goldberger AL. Fractal mechanisms in neural control: human heartbeat and gait dynamics in health and disease. In: Walleczek J, ed. *Self-Organized Biological Dynamics & Nonlinear Control*. Cambridge: Cambridge University Press, 2000: 66-96.
12. Peng CK, Buldyrev S, Havlin S, Simons M, Stanley HE, Goldberger AL. Mosaic organization of DNA nucleotides. *Phys Rev E* 1994; **49** (2): 1685-9.
13. Chatfield C. *The analysis of time series*. Oxford: Chapman & Hall, 1996.
14. Brockwell PJ, Davis RA. *Time Series: Introduction to Times Series and Forecasting*. Berlin: Springer-Verlag, 1996.
15. Voss RF. Random fractal forgeries. In: Earnshaw RA, ed. *Fundamental algorithms for computer graphics* (NATO Series VF17). Berlin: Springer-Verlag, 1985.
16. Landini G. Applications of fractal geometry in pathology. In: Iannaccone PM, Khokha M, eds. *Fractal geometry in biological systems*. Boca Raton: CRC Press, 1995.