

## Lessons from John Graunt

Kenneth J Rothman

John Graunt, 17th century citizen of London, thought of himself as a haberdasher. History knows him, however, as the first epidemiologist and demographer. He made his mark because he grasped the potential in an overlooked data source of his day—the weekly *Bills of Mortality*—and he mined this resource admirably in his only scientific work. His book, published 333 years ago, was a slim edition with a donnish title: *Natural and political observations mentioned in a following index, and made upon the Bills of Mortality*.<sup>1</sup>

With this book Graunt added more to human knowledge than most of us can reasonably aspire to in a full career. Graunt was the first to report, and to document, that more boys than girls are born. He presented one of the first life-tables. He reported the first time-trends for many diseases, taking into account changes in population size. He described new diseases, and noted others that seemed to increase over time only because of changes in classification. He offered the first reasoned estimate of the population of London, demonstrating its rapid growth and showing that most of the growth came from immigration. He proffered epidemiological evidence refuting the theory that the plague spreads by contagion. (He also refuted the notion that plague epidemics are coincident with the reign of a new king.) He showed that the large population decreases in plague years were offset by large increases in births in subsequent years. He showed that physicians have twice as many female as male patients, but that more males than females die. He produced the first hard evidence about the frequencies of various causes of death. And, presaging our present-day paranoia, he tried to allay unwarranted anxiety about risks that were feared far out of proportion to their likelihood of occurrence (panel).

Graunt's scientific attainment is instructive for the ways in which it illustrates principles of epidemiological research and publication that concern us today. Consider his statement of objectives. He noted that the *Bills of Mortality* were published every week, but few who read them made any use of them, apart from gossip about the number of burials. Graunt summarised the weekly bills into tables by season, by year, and by geographic area, and, in his words:

"I did then begin, not onely to examine the Conceits, Opinions, and Conjectures, which upon view of a few scattered *Bills* I had taken up; but did also admit new ones, as I found reason, and occasion from my *Tables*".<sup>1</sup>

Those who have been accused of data dredging can take heart in knowing that the first published epidemiological study was introduced as a search for new

**Panel: Notorious diseases and casualties, representing Graunt's attempt to allay fears about certain risks by showing how few deaths actually occurred (of a total of 229 250) in these categories**

### Table of Notorious Disease.

Apoplex.....	1306
Cut of the Stone.....	0038
Falling Sickness.....	0074
Dead in the Streets....	0243
Gowt.....	0134
Head-Ach.....	0051
Jaundice.....	0998
Lethary.....	0067
Leprosy.....	0006
Lunatique.....	0158
Overlaid, and starved.	0529
Palsy.....	0423
Rupture.....	0201
Stone and Strangury..	0863
Sciatica.....	0005
Sodainly.....	0454

### Table of Casualties.

Bleeding.....	069
Burnt, and Scalded.....	125
Drowned.....	829
Excessive Drinking.....	002
Frighted.....	022
Grief.....	279
Hanged themselves.....	222
Kil'd by several accidents.....	}1021
Murthered.....	0086
Poysoned.....	014
Smothered.....	026
Shot.....	007
Starved.....	051
Vomiting.....	136

hypotheses. Not only did Graunt engage in a so-called fishing expedition, but also he did it with data collected for another purpose. Thus the tradition of capitalising on available data for research purposes is a practice that we can trace back to the beginning of epidemiology.

The importance of the weekly *Bills of Mortality* to Graunt reminds us that epidemiology is largely dependent on the availability of good records. As much as we may think of epidemiological research as being characterised by the high-tech application of data processing and analytical methods, it is not nearly so dependent on technology as many would believe. We could get along fine today, if we had to, with low-tech methods, but not without the existence of systematic records. What exactly were the *Bills of Mortality*? Graunt explained them in what we would now call the methods section of his paper:

"... the rise of keeping these Accompts, was taken from the *Plague*: for the said *Bills* (for ought appears) first began in the said year 1592, being a time of great *Mortality*; and after some disuse, were resumed again in the year 1603, after the great *Plague* then happening likewise. These bills were Printed and Published, not onely every week on *Thursdays*, but also a general Accompt of the whole Year was given in, upon the *Thursday* before *Christmas Day* . . .

When anyone dies, then, either by tolling, or by ringing of a Bell, or by bespeaking of a Grave of the *Sexton*, the same is known to the *Searchers*, corresponding with the said *Sexton*. The *Searchers* hereupon (who are antient Matrons, sworn to their office) repair to the place, where the dead Corps lies, and by view of the same, and by other enquiries, they examine by what *Disease*, or *Casualty* the corps died. Hereupon they make their Report to the *Parish-Clerk*, and he, every *Tuesday* night, carries in an Accompt of all the *Burials*, and *Christnings*, hapning that Week, to the *Clerk* of the *Hall*. On *Wednesday* the general Accompt is made up, and Printed, and on *Thursdays* published and dispersed to the several Families, who will pay four shillings *per Annum* for them".

*Lancet* 1996; **347**: 37–39

Department of Epidemiology, Boston University School of Public Health, and Department of Preventive Medicine, Boston University School of Medicine, Boston, MA, USA (K J Rothman <sup>DrPH</sup>)

Correspondence to: Dr K J Rothman, *Epidemiology*, 1 Newton Executive Park, Newton Lower Falls, MA 02161, USA

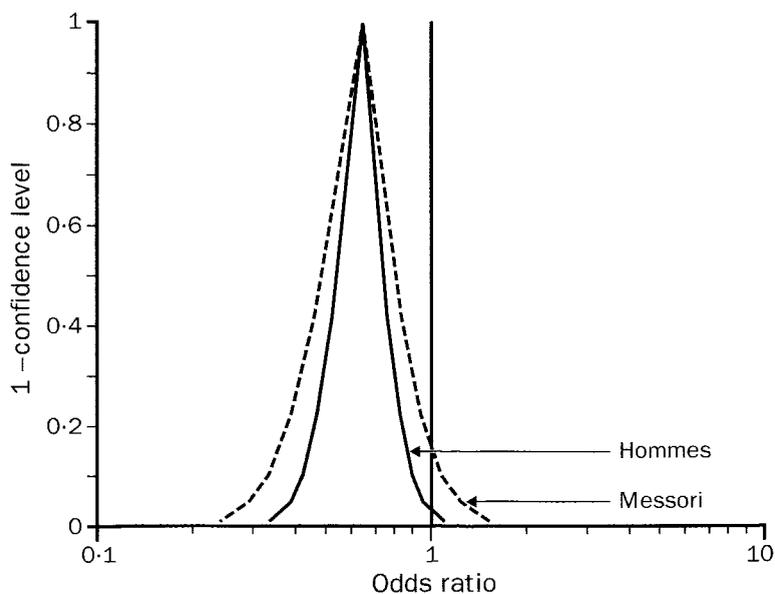


Figure: **Confidence interval functions<sup>7</sup> for Hommes and co-workers<sup>4</sup> meta-analysis and recalculation by Messori et al.<sup>3</sup>**

The figure clearly shows that the two calculations are consistent with respect to a causal effect. (Reproduced with permission of *Epidemiology*).

Graunt's account of the bills set an admirable precedent for scrutinising the origin of epidemiological data. He described the reporting mechanics of the bills, but in an impressively modern spirit he also critically examined errors in the recording and collection of the bills. For example, Graunt knew that plague deaths were of primary interest to those who read the weekly bills, but he found that plague deaths were underascertained. He noted that in years with a large number of plague deaths the non-plague deaths also increased. He inferred that about 20% of plague deaths were mistakenly recorded as deaths in other categories, and on that basis he gave an improved estimate of the effect of the plague on mortality. Thus he considered the difficulty of outcome misclassification, provided an estimate of its magnitude, and used that to correct for the misclassification.

Graunt also speculated on the reasons for misclassifications. Assigning the deaths to the proper categories sometimes called for subjective decisions on the part of the searchers, who had no coroner to help. In his criticism of the data, Graunt weighed the trustworthiness of the field decisions made by the searchers. He knew that even the untrained could recognise a corpse, and could usually determine whether the death was from hanging, drowning, leprosy, or other clear-cut causes. But some causes of death he thought more difficult to identify. With respect to consumption, he claimed that:

“ . . . all dying thereof die so emaciated and lean (their *Ulcers* disappearing upon Death) that the Old-women *Searchers* after the mist of a cup of *Ale*, and the bribe of a two-groat fee, instead of one, given them, cannot tell whether this emaciation, or leanness were from a *Phthisis*, or from an *Hectic Fever*, *Atrophy*, &c or from an *Infection* of the *Spermatick* parts . . . ”<sup>1</sup>

In bluntly describing the searchers as drunken and venal old women, Graunt seems to have cast a jaundiced eye on a crucial part of his data-gathering operation. Rather than excoriate his lack of political correctness, we should admire his fastidious concern for measurement accuracy. How easy it is to submerge such concerns and merely thank the data gatherers in the acknowledgments to one's work. Graunt was not being unkind; he was just being a good scientist.

Graunt's work harbours other lessons for modern day epidemiologists:

(1) He was brief. (No need to elaborate on this statement.)

(2) He made his reasoning clear. He explained his calculations in detail, as, for example, in his estimation of the magnitude of under-reporting of plague deaths.

(3) He subjected his theories to repeated and varied tests. For example, he estimated the population of London to be about 384 000. He derived this figure with five distinct approaches to the estimation, based on methods involving population turnover, births, and housing.

(4) He invited criticism of his work. He modestly asked that readers “correct my *Positions*, and raise others of their own: For herein I have, like a silly Scholeboy, coming to say my Lesson to the World (that Peevish, and Tetchie Master) brought a bundle of Rods wherewith to be whipt, for every mistake I have committed”. Such refreshing modesty, if more prevalent today, might breed greater forbearance among scientists who would prefer only to detail relentlessly the mistakes of others.

(5) Graunt was willing to revise his thinking in the face of his data—he admitted to developing some new conjectures on review of his tables. Such an admission might sink a proposal or a submitted manuscript today, when we are likely to be advised that we should not study a question that was not specified in gory detail before we began to collect our data. Graunt reported that there are more male than female births; he did not say that his observation should only be regarded as a “hypothesis generating observation” and that another study, a “hypothesis testing study”, was needed to address the question definitively.

Today we seem burdened by regimented thinking that amounts to science by recipe. Two factors motivate this regimentation. One is fear of false leads, that unanticipated findings could be merely the play of chance. And well they could. But without probing possible leads, we might miss important results lurking in the data, so why should we be reluctant to probe? The second factor is the possibility of deceptive analyses that distort a finding by gerrymandering category boundaries. When someone asks you to specify category boundaries in advance, and tells you that they should not be modified, the hidden message is “we don't trust you to describe the data honestly and competently”. Your initially chosen boundaries might make no sense in the light of the actual data. It could happen that the category boundaries do affect the results; in one published example a relative risk estimate changed from 0.8 to 3.3, depending on whether the mean or the median was used to dichotomise exposure to butadiene.<sup>2</sup> But when that happens, that is the finding. In such a situation, reporting only a single value based on a prespecified cutoff fails to describe the data fully.

(6) The ultimate mechanical interpretation of data is the enduring vexation of statistical significance testing. Graunt had no knowledge of, nor any need of, statistical significance testing. Our present-day preoccupation with statistical significance stems from a tropism to have clear-cut, black-and-white, formula-driven answers to the complicated questions that we study. Huge and informative bodies of data have been debased into dichotomous categories because so many have been trained to ask only, “is it significant?”.

Even worse, significance testing often points us in the wrong direction. Consider as an example the following published comment about the efficacy of subcutaneous heparin in treating deep vein thrombosis: "The recent paper by Hommes and colleagues reports a meta-analysis of six randomised trials comparing subcutaneous heparin with continuous intravenous heparin for the initial treatment of deep vein thrombosis . . . The result of our calculation was an odds ratio of 0.61 (95% CI, 0.298 to 1.251;  $p > 0.05$ ); this figure differs greatly from the value reported by Hommes and associates (odds ratio, 0.62; 95% CI, 0.39 to 0.98;  $p < 0.05$ ) . . . Based on our recalculation of the overall odds ratio, we concluded that subcutaneous heparin is *not* more effective than intravenous heparin, exactly the opposite to that of Hommes and colleagues . . ." <sup>3</sup> Note that the authors of this letter estimated an effect that was even stronger than the originally reported result. <sup>4</sup> Yet they took this stronger effect estimate to be evidence against there being any effect at all, apparently because it was not statistically significant.

Instead of thinking dichotomously in terms of what is or is not statistically significant, we are much better off to rely, as Graunt did, on estimation. The confidence interval is the usual way to measure random error, although we need to avoid a common trap in their use: because the intervals involve setting an arbitrary confidence level, they can be used to make a statement equivalent to assessing statistical significance. But, as Poole <sup>5</sup> has emphasised, a confidence interval that is used merely to determine whether one of the bounds crosses the null point or not is misused. The boundaries of a specific confidence interval are not the regions in which our attention should be focused in the first place. Instead, it is two quantitative ideas—the general location of the interval, and its breadth—that one should take from the interval.

It is best to think of a specific confidence interval as a surrogate for a curve that represents all possible confidence intervals. One way to generate such a curve is to calculate the p-value for every possible hypothesis of

the parameter of interest, the so-called p-value function. This curve includes the ordinary null p-value as one point, along with all the non-null p-values. From it one can also read all possible confidence intervals, and so the function is sometimes called the confidence interval function. <sup>6</sup>

The figure, published in an earlier critique, <sup>7</sup> shows the two p-value functions for the two sets of findings for subcutaneous heparin. The two estimates are clearly essentially the same, apart from some small differences in precision. These differences stem from one of the two sets of investigators using the wrong variance formula in their meta-analysis calculations. <sup>8</sup> It hardly matters which one used the wrong formula—no sensible person should draw different conclusions from these two curves. This misinterpretation, and myriad others like it, are the result of reliance on significance testing, a peculiarly modern affliction in science. We can, and should, avoid statistical significance testing. John Graunt's influence can persist by providing inspiration to think clearly, openly, and modestly, and to avoid the harmful habits that we have acquired.

This paper is abridged from the 1995 John Cassel Memorial Lecture, presented at the annual meeting of the Society for Epidemiologic Research, Snowbird, Utah, June 22, 1995. I especially thank Philip Cole, Elaine Ron, Aaron Cohen, and Manning Feinlieb for helpful comments.

#### References

- 1 Graunt J. Natural and political observations mentioned in a following index, and made upon the Bills of Mortality. Facsimile edition. New York: Arno Press, 1975.
- 2 Cole P, Delzell E, Acquavella J. Exposure to butadiene and lymphatic and hematopoietic cancer. *Epidemiology* 1993; **4**: 96–103.
- 3 Messori A, Soraccaro G, Marini N. Calculation errors in meta-analysis. *Ann Intern Med* 1993; **118**: 77–78.
- 4 Hommes DW, Bura A, Mazzolai L, Bueller HR, tenCate JW. Subcutaneous heparin compared with continuous intravenous heparin administration in the initial treatment of deep vein thrombosis. *Ann Intern Med* 1992; **116**: 279–84.
- 5 Poole C. Beyond the confidence interval. *Am J Public Health* 1987; **77**: 195–99.
- 6 Sullivan KM, Foster DA. Use of the confidence interval function. *Epidemiology* 1990; **1**: 39–42.
- 7 Rothman KJ, Lanes S, Robins J. Causal inference. *Epidemiology* 1993; **4**: 555–56.
- 8 Goodman SN. Editor's response to Messori et al. *Ann Intern Med* 1993; **118**: 78.