International Statistical Review, 48 (1980) 169–195 Longman Group Limited/Printed in Great Britain

Representative Sampling, IV: the History of the Concept in Statistics, 1895–1939

William Kruskal and Frederick Mosteller

University of Chicago, and Harvard University

Summary

We trace ideas of representative sampling in the history of statistics, starting with Kiaer's (1895) advocacy of what he called the representative method and ending just before World War II. A major theme is the slow introduction of lucid probabilistic ideas into the domain of statistical surveys, a process still far from complete. Other themes include (1) the early debate between statisticians advocating sampling and those insisting on attempted full coverage, (2) early American discussions of representativeness – generally superficial – and their use by Kiaer as supporting arguments, and (3) the experience of Gini and Galvani (1929), and the roles of Bowley, Neyman, and others.

1 Introduction

Our three prior papers on representative sampling, Kruskal and Mosteller (1979a, 1979b, 1979c), dealt respectively with the concept in non-scientific writing, in scientific writing outside statistics, and in the statistical literature. Although those three papers included sporadic historical comments, we now present a more systematic exposition of the history of the idea of representative sampling, especially within statistics itself.

At the outset, we call attention to a problem that we faced as authors and that you face as reader. We review both words and ideas: it would be pedantic to become over concerned with particular words or expressions except insofar as they are intertwined with ideas. Yet the ideas connected with words like 'representative sample' are so basic and range so widely that a full historical account would be tantamount to writing a history of nearly all statistics – indeed, of a large segment of philosophy of science. To keep our efforts within manageable limits, we concentrate here on the history of usage within statistics of expressions like 'representative sample' and 'representative method'. We digress slightly from time to time. By and large, we treat the period after 1894 and before 1940; our third paper, on current statistical usage, dealt mainly with material published after 1940, but we are not rigid about chronological boundaries.

To help our recall of the nine kinds of usage for 'representative sampling' that we described and exemplified in our earlier papers, we provide drawings with captions. Most, but not all, of these turned up in our exploration of the pre-1940 statistical literature.

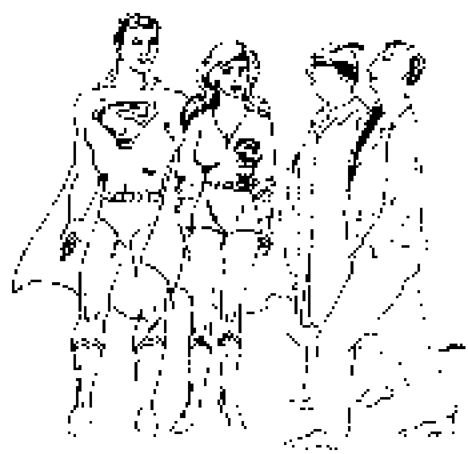
Our historical treatment begins with a seminal speech by Kiaer (1895), at the Berne meeting of the International Statistical Institute, and we then use reports of later ISI meetings to provide both structure and material for discussion. We treat connections between Kiaer and American statisticians, most notably Carroll D. Wright. The latest ISI meeting that we discuss is that of Rome in 1925; its discussions led to Jerzy Neyman's fundamental Royal Statistical Society paper of 1934.

Perhaps the major theme of our exposition is the painfully slow – in retrospect – way that

Meaning 1. *General, usually unjustified acclaim for data: The emperor's new clothes.*

Meaning 2. *Absence of selective forces: Justice balancing the scales.*

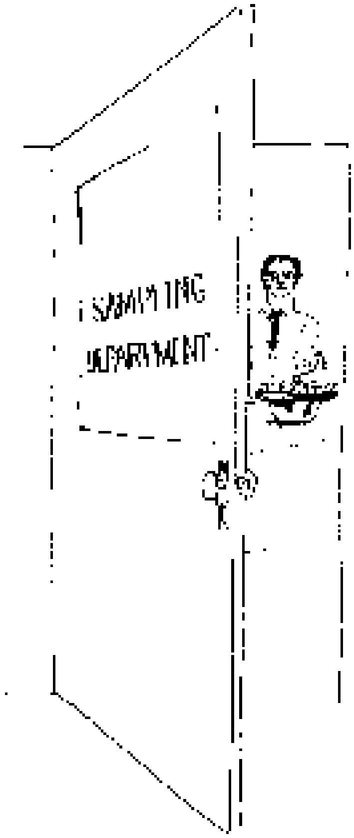
Meaning 3. *Miniature of the population: Model train set.*



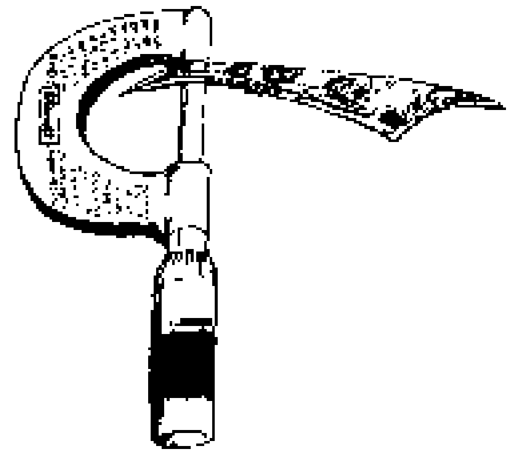
Meaning 4. *Typical or ideal case: Superman and Superwoman and Average man and Average woman.*

Meaning 5. *Coverage of the population: Noah's Ark*

Meaning 6. *Vague term to be made precise: A hatching egg of unknown species.*



Meaning 7. *Some specific sampling method: The Sampling Department in action.*



Meaning 8. *Permitting good estimation: Micrometer*

Meaning 9. *Good enough for a particular purpose. Boy fishing with worm and safety pin.*

precise ideas of probability entered the theory and practice of surveys.[1] Progress was halting and dogged by confusions and misunderstandings. Little interaction occurred between social and economic survey takers, on the one hand, and statistical analysts in the natural sciences, on the other. Karl Pearson might have formed a bridge between the two statistical worlds, but he did not. We think that it was the new investigations of the foundations of probability in the 30's, together with Neyman's influence following hard on Bowley's, that brought us to the modern era.

It is dangerous, as well as discourteous to our intellectual forbears, to crow about our present knowledge. In the first place, there are still many gaps, and there is still widespread ignorance and confusion about representativeness, probability, and the need for stringent standards in the design and analysis of surveys. Second, inevitably our intellectual descendants, fifty to eighty years from now, will look back and wonder at our failures to understand. In that sense of humility, we embark on our historical exposition.

2 The beginning

To our knowledge, the first analytical[2] use of a term like 'representative sampling' occurred at the 1895 Berne meeting of the International Statistical Institute and was published (Kiaer 1895–1896) in the report of that meeting. Anders Nicolai Kiaer, Director of the Norwegian Central Bureau of Statistics, then in his late fifties, had been fascinated by the notion of obtaining useful national information from social and economic data that did not cover all individuals, households, or other units of the entire nation (or other region of interest). His 1895 paper, published in 1896, is entitled 'Observations and experiences with representative surveys' (Observations et expériences concernant des dénombrements représentatifs). Kiaer was sharply opposed by many of the leaders in government statistics who formed the core of ISI and who argued for full enumeration, but Kiaer continued to press, both in and out of the Institute, for what he came to call most frequently the 'representative method', that is, for sampling. Thirty years later, by the time of the 1925 ISI meeting in Rome and after Kiaer's death, representative sampling had found an accepted, honorable place among statisticians. Sampling had become respectable, if not yet well understood.

Of course sampling, in one sense or another, must be as old as mankind. One arrives at conclusions about fish, trees, lions, wood nymphs, and fire on the basis of relatively small unstructured samples that come to hand in daily life. All of us continue such informal inferences most of the time. Yet that is a far cry from the self-conscious ideas of sampling and of inferences from samples that have gradually and erratically grown over the last two hundred years.[3]

In geology, agriculture, and physics, for example, one simply cannot know the whole population, so samples are inevitable and essential. One main stream leading to modern statistics

---

[1] We recognize many difficulties in this judgment of slowness. First, such a judgment is a cliché of intellectual history; as William Jaffé said, '... how slow is the process of growing from generation to generation toward clarity, whether it be in economics or in any other science' (1965) (p. 220).

Second, one must ask: slow compared to what? In the history of statistics, some major advances have been remarkably fast, for example, (1) randomization in design and (2) the Neyman–Pearson interpretation of confidence intervals. These were relatively specific advances, to be sure, and came directly from one or two exceptionally vigorous minds.

Third, quantitative statements about slowness inevitably provoke statisticians trying to study history because of the statements' imprecision. How might one possibly measure speed of innovation with care and make quantitative comparisons? Of course others have given thought to these questions, and we have some tentative ideas ourselves, but they are for other places and other times.

[2] As we shall see later, Carroll D. Wright used such an expression earlier, but in a shallower and less influential way.

[3] An ancient method selects subsets of a population by the use of lots, that is, by some kind of physical randomization that picks juries to hear a dispute, chooses soldiers to be killed as punishment for the entire regiment, allocates land or other property, etc. To trace this kind of randomization back to Assyria and the Old Testament Jews see Fienberg (1971) (p. 255). Lichtenstein and Rabinowitz (1971–72) and Ehrenberg (1927). Rabinovitch (1973) and its review by Zabell (1976) provide useful material. For judicial and legislative uses of lots in ancient Greece and

rises from problems in the natural sciences and was deepened by such major figures as Laplace and Gauss. These authors typically treat a sample as it if were random (*i.e.*, composed of identically distributed independent observations), and frequently make a parametric assumption (*e.g.*, normality). In such a setting, one needs the concept of representative sampling less than in social surveys. The concept, however, might well have arisen in geological or astronomical sampling, but we have not run across such an origin.

Social and economic statistics themselves have a long history of sample studies, in the sense of partial coverage of the relevant population. Then, in the 19th century, with its growth of national statistical systems and the contemporary but separate development of sociological studies, one sees two almost polar movements. The first pushes for full coverage, for the complete census, the second cultivates case studies – sometimes called monography – to permit insight and hypothesis formation by intensive investigation of a few families, a single city, one aboriginal tribe, or, as for Freud, a handful of extreme neurotics in one particular milieu. See Lazarsfeld (1961); Lécuyer and Oberschall (1978); and Kruskal (1978b).

Full coverage and the case study idea need not be in conflict. For example, a case study of a town might include a census of everyone reachable. As Hill (1966) said,

> Readers of the early volumes of the *Journal of the Royal Statistical Society* will be struck by . . . [this:] In their surveys of . . . housing, education, or health. . . . in the 1830's, it was the aim of the pioneers of that time to study and enumerate *every* member of the community with which they were concerned – the town in Lancashire, the borough of East London, the country village, whatever it may have been. That aim was frequently brought to naught by the very weight of the task . . . [,by] collection of the data . . . [and] the statistical analysis that the results demanded. In contrast, the worker today would (or should) instinctively reflect on the possibility of . . . sampling. (P. 14.)

We strike one terminological note before we turn to the sequence of Kiaer's papers, for terminology tends to be tangled, with various writers using the same expression in different senses, and with the same writers using different expressions for the same idea. An intentional investigation of fewer than all units in a population, we might today well call sampling. In the late 19th century such terms as 'partial investigation' and 'indirect methods' seem to have been popular. For discussion of the latter see Westergaard (1932) (p. 263).

To sum up, the period 1895 to 1925 marked a transition for social and governmental statistics from heavy reliance on complete coverage to substantial additional use of designed partial investigations. Anders Kiaer and others promoted and used such partial investigations. Confusion reigned about the meaning of representative sampling.

Rome see Staveley (1972, index entries are under 'sortition'). For Roman decimation, see Fiebiger (1901); we also note the reference to this practice by Shakespeare in *Timon of Athens*, Act 5, scene 4.

> . . . take thou the destin'd tenth,
> And by the hazard of the spotted die
> Let die the spotted.

Ancient randomization was, of course, also used for divination and for games. Little of the ancient literature that we have seen gives self-conscious discussion of the role of randomization to produce an equitable result or a representative sample in one or another sense. Nor have we found in the ancient literature any analysis of sampling for empirical study. We may, of course, have missed or failed to recognize relevant ancient materials.

As we have said, we forbear discussion of the use of sampling *per se*, a practice that can hardly have had a sharp beginning. See Stephan (1948). The idea of sampling and letting the sample stand for the entirety arises widely, and we cite here, by way of example, only a glittering selection of three works about sampling connected with gold. S. Stigler (1977) describes the trial of the Pyx, an English test of coinage going back perhaps as far as the twelfth century. Rickard (1904) has a discerning discussion of sampling gold mines; without reference to the statistical literature, he has many wise remarks about sampling from a highly heterogeneous, skew population under circumstances of distrust and frequent sharp practice between buyer and seller. Dewey (1913) analyzes ways of sampling gold bullion.

3 Kiaer and the International Statistical Institute[a]

In between full coverage and the case study came the proposal for self-conscious sampling, the representative sample, under the initial leadership of Kiaer. What seems new in 1895 is the fresh concern about the relationship between sample and population, a concern that at first has almost nothing to do with probability.

The newly named representative method met searching criticism from some of the old guard adherents of full coverage. Perhaps the most vocal was Georg von Mayr, who said, for example, in the discussion after Kiaer (1895–96) that, in our translation and paraphrase,

> ... I regard as most dangerous the point of view found in his [Kiaer's] work. I understand that representative samples can have some value, but it is a value restricted to terrain already illuminated by full coverage. One cannot replace by calculation the real observation of facts. A sample provides statistics for the units actually observed, but not true statistics for the entire terrain.
>
> It is especially dangerous to propose representative sampling in the midst of an assembly of statisticians. Perhaps for legislative or administrative goals sampling may have uses – but one

---

[a] We know of several briefer treatments of the ISI meetings at which the 'representative method' was discussed. Two, Zahn (1934) (p. 64–65) and Nixon (1960) (p. 88), are histories of the ISI after 50 and 75 years respectively of activity. The third is a deeper discussion by Dalenius (1953). Seng (1951) gives a lengthy treatment of the material from a viewpoint rather different from ours. Kenessey (1963) provides a fresh presentation with special attention to representative sampling in eastern Europe; he includes (p. 32–3) a translation into English of a highly interesting 1894 letter from Kiaer to J. Körösy.

Zoltan Kenessey and Stephen M. Stigler point out in correspondence that the opposition to Kiaer's ideas about sampling, an opposition led by von Mayr and other official statisticians, may well have been based on their view of the then-recent history of social statistics. High standards of objectivity and completeness had not been long established, and were associated with full censuses; a shift to sampling might understandably have been viewed as retrogression. Kenessey also suggests that the official statisticians may have viewed sampling as a threat to their budgets and bureaucratic establishments. We hope that others will try to find documentary evidence about these plausible conjectures.

Stephan (1948) has a basic article on the history of sampling, and Chapter 2 of Dalenius's 1957 book contains valuable historical material. We also cite with appreciation the 1976 paper by Hansen and Madow.

Early work in Russia on the theory and practice of sample surveys is relevant, and we note two papers by Zarkovic (1956, 1962). The 1956 paper briefly describes (p. 337) a 1921 article by A. Gurev that 'shows that the ideas popularized by Kiaer were used in Russian theory and practice in the second half of the nineteenth century, i.e., much earlier than when Kiaer started ...'. Time and language limitations have kept us from examining adequately this early Russian work.

We have, however, examined Kaufmann's 1913 book, which describes (p. 362–374) Russian work on sampling for social and economic surveys. A remark on p. 427 of Jensen (1926b) sent us to Kaufmann's book: Jensen says that representative investigations in Russia are 'easily accessible in Al. Kaufmann's instructive work'. Kenessey (1963) also refers to Kaufmann and discusses early Russian sampling.

According to Kaufmann, Russian ideas of representative sampling derived from Kiaer's work in the 90's, so there seems to be something of a contradiction between Gurev and Kaufmann, possibly based on ambiguity of meaning. We hope that others will clarify this historical puzzle.

Kaufmann, who was at the University of St. Petersburg according to his title page, describes Bowley's work, makes brief mention of von Bortkiewicz's ideas, and treats the early and sometimes neglected sampling theory of A. A. Chuprov (transliterated Tschuprow, Tchuprow in German).

Kenessey points out that statistical issues related to representativeness played a role in the debates that formed an overture to the Russian revolution. Chapter 2 of Lenin's 1899 book, *The Development of Capitalism in Russia*, is an analysis of the Zemstvo statistics for the Russian countryside. (The Zemstvos were rural government units.) Lenin has two major points there that are relevant to our exposition. First, he takes the Zemstvo statisticians to task for looking at averages only rather than at the variability whose analysis – says Lenin – shows the formation and activity of economic classes in the Marxist sense. See p. 73 for example. Second, the Zemstvo household budget statistics show too rosy a picture (p. 151) because the very concept of a budget '... is not easily found among the poor'. Since the early history of the Russian revolution turns to a large extent on the treatment of the wealthier peasants, the statistical points raised by Lenin may well have had a significant effect on Soviet policy.

We suppose that the Zemstvo statisticians must have had responses to Lenin's criticisms, but we do not know that literature. Lenin's scathing reply to one rebuttal is called 'Uncritical criticism' (Lenin 1956, p. 617–640, originally published 1900).

For a detailed description in English of the Zemstvo statistics, see Kaufmann's chapter on Russia in Koren (1918).

must never forget that it cannot replace a complete survey. It is necessary to add that there is among us these days a current in the minds of mathematicians that would, in many ways, have us calculate rather than observe. We must remain firm and say: no calculations when observations can be made. (P. XCIV.)

Ringing words! 'Pas de calcul là où l'observation peut être faite.' An excellent brief statement about the von Mayr position of the need for completeness and how it was overturned was given by Anderson (1934) (p. 368).

At this remove we cannot easily understand von Mayr's references when he talks of calculation and mathematics. Kiaer himself says hardly a word about probability, nor of mathematics in any real sense. He only takes ratios and averages in his representative samples as holding approximately for the population, and he provides what evidence he can that the approach works.

What exactly did Kiaer have in mind by representative sampling? First, he thought of social and economic surveys in which one could begin by choosing districts, towns, parts of cities, streets, *etc.*, to be followed by systematic, rather than probabilistic, choice of units (houses, families, individuals). Second, he insisted on substantial sample sizes at all levels of such a selection process. Third, he emphasized the need for spreading out the sample in a variety of ways, primarily geographically, but in other ways as well. For example, if a sample had a deficiency of cattle farmers, he would add more of them.

He returns often to the idea of coverage in various modes. For example he says that 'we should be careful not only to visit houses that are average from a social point of view, but also houses representing the different social and economic conditions that are found in the community'. (1895/6, p. 179.)

Kiaer thought of his representative sample as an approximate miniature of the population. He repeats that idea time after time and stresses the importance of comparing sample results with known census results wherever possible: age, marital status, occupation, *etc.* The advantage of the sample, of course, is that one can get far more detail, detail that would be impractically expensive for a census: that is, the sample provides information on many more characteristics than the census.

A year or two later, Kiaer (1897) expressed himself at greater length, but in Norwegian. The article has recently been translated, and for readers whose major language is English, the translation presents an excellent, leisurely picture of Kiaer's thought.

One novelty of this paper is brief mention (p. 39 in the translation) of possibly drawing the sample by lot. He does not actually do that, however, but chooses certain ages and certain initial letters of surnames: '. . . [these] seem to be related to the persons from amongst whom the sample should be selected in a haphazard and random way, so that a sample selected in this manner would turn out . . . [as if it had] been selected through the drawing of lots in order to avoid, in the most stringent manner, any procedure that would give preference to persons in certain occupations or belonging to particular social strata'. (It may turn out that use of initial surname letter does introduce unfortunate selection bias. See Jensen (1926b) (p. 429–430) for an example in a Mannheim study: Jensen in addition refers there to Kiaer's use of selected ages and consequent difficulty because of age rounding.)

Kiaer also suggests (page 39 in the translation) differential sampling rates in strata with subsequent weighting back for the analysis. The motivation given is obscure, and of course any such approach — however praiseworthy — spoils the simplest miniature idea.

He points out the prior use of sampling ('partial investigations') and gives examples that include the following:

Grain harvest. ('[It] . . . would simply be impossible [to take] . . . reports from each and every agricultural holding.' P. 54.)

Numbers of letters mailed between countries ('. . estimated by . . . a count of the volume at certain weeks during the year'. P. 54.)

Geological surveys. ("[From surveys] . . . in scattered locations, inferences are drawn . . . and a picture is formed of the whole area.' P. 37.)

Yet he proudly feels (p. 48) that his surveys in Norway were the *first* properly representative ones, in that they (1) were planned with representativeness as the goal, and (2) checked insofar as possible against census results. At the same time, he recognizes imperfections in those surveys and hopes for improvement.

He likes the two-barrel analogy: think of a large barrel and a small barrel, both of mixed contents. If a number of samples from each show their contents to have practically the same composition, then further analysis of only the small barrel permits inferences about the large one (p. 51).

St. Petersburg, 1897 Kiaer (1899a) repeats his themes at the 1897 ISI St. Petersburg meeting. He sharpened his concept a bit and brought out the miniature idea more clearly.

By representative investigation I understand a partial exploration with observations on a large number of scattered localities, distributed over the whole territory so that they form a miniature of that whole. The localities are not chosen arbitrarily, but according to a rational grouping based on census results; and the results should be controlled by comparison with those censuses. (P. 180; this is a free translation.)

He also draws a further contrast between representative sampling and case studies (i.e., typical cases, monographic studies): a case study of a particular type, or even several such case studies, do not give a 'true miniature of the totality' with its 'variety from case to case' (p. 181).

Kiaer participated in the 1897 Stockholm Conference of Scandinavian statisticians; see Dalenius (1957) (p. 30). The Conference agreed that representative survey methods were useful, especially in social statistics, but urged that complete statistical surveys be preserved as far as possible. We can sympathize with another recommendation that, when a representative survey is carried out, a complete report should be made of how the sample was selected.

In 1899 Kiaer (1899b) published in von Mayr's journal a statement of his position. We find nothing in it that does not appear in his other publications except for his (apparently) first citation of American ideas and experiences towards the use of representative sampling. We return to this point in Section 4.

Budapest, 1901 At the next meeting, in Budapest, Kiaer (1903) returns to his themes, with some new emphases. First, Kiaer stresses the need for spreading out the geographical units over the country. He adds (p. 66) that if the units are gathered haphazardly (au hasard) so that many fall in one area and few in another, then one has a partial investigation but not a *representative* one. Here 'au hasard' does not carry any probabilistic connotation.

To check representativeness without a relevant census for comparison, he proposes to 'divide the investigation into two or three separate parts, each itself a representative study. If the parts give similar results, one cannot deny the validity thus gained' (p. 68). Kiaer does not note the circularity that arises because we cannot know that the separate parts are themselves representative by his standards. Nonetheless, we see here an early notion of replication and subsampling.

In the discussion after Kiaer's talk, von Bortkiewicz made a noteworthy contribution. He points out the vagueness in Kiaer's requirement of approximate agreement between the sample and the population in order that the term 'representative' be merited, and Bortkiewicz suggests, in effect, a significance test. He says (p. 72), 'Can the observed difference be considered as random [due au jeu du hasard, aux causes fortuites ou accidentelles]? The formulas needed have been deduced by Poisson for analogous cases; they permit easy calculation of how much difference between two ratios or coefficients can be attained by accidental causes. If the observed difference in Kiaer's case between population and sample lies outside those limits, then representativeness fails. Otherwise it may hold. I have applied this approach to Kiaer's examples in his *Allgemeines Statistisches Archiv* article, and the observed differences greatly exceed the limits of accidental variation. Thus Kiaer's great skill [l'extrême prudence et la grande sagacité] in selecting his samples does not result in the exactitude that one theoretically requires. Hence I reject the term "representative method" for those examples."

The Bortkiewicz quotation is a fascinating one. To begin with, it is the first serious introduction of probability in the sequence of ISI papers on the representative method; second, it applies a simple probability model to a clearly *non*probabilistic sample; and third, it uses a significance test as a go–no go device. The last two questionable procedures have been with us at least since Laplace. For a discussion and for bibliography see Kruskal (1978a).

Kiaer (1903) (p. 77–78) replies to some minor objections, but we have nowhere found his reply to Bortkiewicz's attack based on statistical significance. The two statisticians, we suspect, were on quite different wavelengths. Von Bortkiewicz does not give enough information for us to review his methods. It seems likely that the variability he assesses would be associated with random sampling rather than with sampling of clusters or chunks. Thus he might compute variances much too small for sampling methods analogous to Kiaer's.

Kiaer also notes that he 'prefers a small number of careful observations, carried out with great care, to a large number of superficial observations made superficially on a large scale'. (P. 78.) That argument is, of course, now conventional wisdom in defense of sampling as opposed to full coverage; in 1901 the argument was fresh.

The American connection arises again, as Kiaer quotes with pride from a letter written by Carroll D. Wright, founder of the now called Bureau of Labor Statistics. We shall return to this after describing events at the 1903 ISI meeting in Berlin.

Berlin, 1903. The report of that 1903 meeting contains a recapitulative statement by Kiaer (1905) with little beyond his earlier materials. One innovative remark, however, is.

> It is true that the results of the representative method are only approximations, but that is the case for all statistics even when it is a question of drawing conclusions from complete observations. The differences between the ratios in a representative sample and in the population are scarcely larger than those between the results from one full census to another. (P. 121–122.)

That is the first explicit statement we noticed in this sequence of discussions about variability over time of the population itself. The remark is quantitatively dogmatic, and without evidence, although some might have been given. More important, the remark could lead us to a familiar methodological discussion: do we wish to take a population as fixed, or as itself arising from an underlying stochastic process?

The discussants ran through familiar criticisms and defenses of representative sampling, but we call special attention to the contribution of Lucien March on p. 129–131. March takes up, in effect, the probabilistic theme begun by Bortkiewicz two years earlier (although March does not refer to Bortkiewicz).

First, March says that the representative method has been used for a long time. Its most illustrious example is the estimation of the French population in the 18th century by counting populations in scattered districts, computing the ratio of population to number of births for those districts, and then applying it to the whole nation (for which the number of births is supposed known). Similar ratio approaches had been used earlier, but with the number of households (hearths = feux) playing the role of the number of births above. More recently, continues March, the 1891–93 salary study of French workers had been based on a sampling of a fifth of the population. Its accuracy was checked by comparisons with older samplings and with administrative records.

(Whether Kiaer agreed that these examples were truly examples of the representative method is not clear. No objection is recorded in the *Bulletin*. Indeed, it would have been difficult and even graceless for Kiaer to fuss or fret, for March was providing both support and – a bit later – a compromise resolution.)

As to the population estimates, continued March, Laplace himself computed the uncertainty of the estimate ("l'erreur à craindre") on the assumption that the districts observed had been drawn at random. 'The formulas that this illustrious savant was the first to give permit no doubt of the method, providing the hypothesis of randomness is verified. In verifying that hypothesis resides the main difficulty.' (P. 130.) Cochran (1978) reviews Laplace's analysis and finds in it a mistake or two, though the mistakes do not seriously affect the conclusions. Other reviews are by Sheynin (1978) (p. 158–161), by Heyde and Seneta (1977) (p. 23), Chang (1976) and much earlier, by Pearson (1928) (p. 166–174); (1978) (p. 462–465).

March uses 'au hasard' to mean at random in the modern technical sense, at least so it seems at first in the context of reference to Laplace. (As we shall see in a moment, March really means probability sampling!) Other discussants in this period sometimes used 'au hasard' to mean haphazardly in a perjorative sense, for example, Kiaer on p. 125, but usage appears erratic, even by the same writer.

Then March says that no one has noticed the need for randomness if the representative method is to hold. But to take a random sample ('prendre au hasard') may have different meanings. For example, it may mean simple random sampling without replacement from a finite population, or it may mean cluster sampling in which the *clusters* are chosen as a simple random sample without replacement. (March describes all this in terms of black and white balls in boxes.)

There are further topics in March's exposition, but he misses the chance to suggest stratification. When the French population estimates were made, he continues, districts were chosen to be well distributed ('bien répartis') geographically; similarly Kiaer requires that the geographical units be 'réparties regulièrement'. But *ipso facto* the sampling is not probabilistic and an analysis based on that hypothesis may be false!

Not only that, other traps may spring. 'Laplace did not suspect . . . that . . . the natality ratio might vary with population size. Natality was probably higher in the smaller towns than in the larger, but one would be tempted to choose smaller towns for economy; consequently, natality would be overestimated and population underestimated. Calculations may be exact, but the hypothesis on which they are based need not hold.' (P. 131.)

March continues with an appeal to include in survey expositions detail about the method of sampling so that the reader can understand and criticize the procedure. That appeal is timeless, and still heeded far too seldom

Thus we see probabilistic ideas gradually entering, not too clearly (we have ventured to interpret March's statement a bit) and not too persuasively. March[5] is important, not only for the

[5] We were surprised to find in other writings of Lucien March, writings published during his long and productive life, somewhat cautious, even negative, views about the role of probabilistic thinking and terminology in statistics. As Michel Huber says in a detailed, analytic obituary,

> March argued for avoiding the introduction of the idea of probability. For him, statistical theory should find an adequate foundation in the principle of compensation that justifies the use of averages. Returning to, and developing, ideas expressed by Bienaymé around 1850, March wished to base the study of statistical distributions on the fluctuation of averages of series chosen randomly in an ensemble. Huber (1933) (p. 277).

This exposition is not fully clear to us, nor is March's own discussion in articles cited by Huber. Part, at least, of March's concern about probabilistic ideas arises from superficial use of conventional probability models without adequate thought about the particular relevant scientific context. For example, see March (1918) (p. 291).

probabilistic flavor of his discussion, but also because he forms in his person a common actor at both the 1903 Berlin meeting and the 1925 Rome ISI meeting. Representative sampling mysteriously dropped out of sight – at least in ISI discussions we have seen – during those 22 years (despite an understanding at the 1903 meeting that there would be a further report in 1907). It could not have been just World War I, nor could it have been Kiaer's inactivity or incapacity – he published through 1919, the year of his death. What was it about la Belle Époque that created a representative hiatus? We do not know.

Summary. To sum up, Kiaer pressed his ideas at the ISI meetings in Berne, St. Petersburg, Budapest, and Berlin and in other writings and he used the method. He had some opposition and some support. He sharpened his descriptions during this period. Von Bortkiewicz introduced and applied a significance test for representativeness, but since Kiaer had no probabilistic model, the test had to be 'as if' he had a specific one. March widened the discussion and developed the idea of probability sampling.*

It strikes us as ironic, although perfectly understandable, that March should be a leader in the introduction of probability models for sampling and at the same time an advocate of caution, even skepticism, about the use of such models.

In March (1912), another article cited by Huber, the criticism of probability again seems to be a warning against taking probability models too literally. March discusses (p. 272–275) the probable error as a measure of the variability of a statistic, and he instructs us not to forget that this procedure has an element of artifice and convention.

> This use of probable error implies that the basic conditions of the observations are analogues to those assumed by probability theory for chance events. But we know that those conditions are not satisfied. Nonetheless, the conventional determination of precision, the probable error, has value because as sample size grows the basic conditions (like those of probability theory) become less and less essential. One should therefore not deal with small samples when one determines indices of precision because the meaning of those indices is uncertain.

From context it appears that the main basic condition is that of normality. On p. 273 March describes in effect 50 per cent confidence intervals (plus or minus one probable error) but of course without clarifying at all the meaning of the interval.

Perhaps March's most extensive discussion of his reservations about probability models appeared in his 1909 ISI article. He makes it clear that probabilistic and other mathematical models are important and useful, but that we must always keep in mind that they are approximations, often with a strong element of convention about them.

Thus Huber's description of March's cautionary view seems to us exaggerated, although it might well be that March's view grew more extreme towards the end of his life and that Huber's summary was based in part on conversations.

* We have asked ourselves when physical randomization (slips of cardboard in a box, balls in an urn, etc.) was first self-consciously used to pick samples for surveys with identified units. Thus we exclude ancient casting of lots to attain equity, we exclude coin and needle throws to illustrate or investigate probability calculations, we exclude simulation studies for distribution theory (like Student's early empirical work in connection with the t distribution), and we exclude independence of observations by assumption as in the classic work by Laplace and Gauss.

These exclusions leave few pure examples; in fact, we have found no fully documented one. One example is described by Jensen (1926b) (p. 387): Norwegian sampling in 1911 and 1921 by mixing slips of cardboard – but without real documentation. Stephan (1948) (p. 21) cites 1906 sampling by Bowley, who used tabled numbers to provide the sample units – but that is not strictly physical randomization. The best documented and the earliest use of physical randomization for a sample survey that we have seen relates to a study of apartments, about 1910, in Göteborg by Edin (1912). Dalenius (1957) (p. 40) translates some of the discussion.

> ... [T]he addresses of all real estates containing such apartments were noted, together with their total, on small slips (cards). These slips were mixed in an urn and taken out one by one, whereupon the notes were transferred to district lists, one for each investigator, until a total of small apartments was obtained slightly exceeding 1/5 of the total number of such apartments which existed in 1905.
>
> *The main object of the procedure here described has naturally been that no one could possibly have the slightest reason for saying that the sample was biased, or on the whole, for whatever reason, that preferably worse apartments have been included* in the survey. As far as I can discover, the method used is the only one which can be called representative, in the strict meaning of the word – something which is not valid when that form of sample is used where a few suburbs are surveyed more or less completely.

It is surprising to us that early examples of physical randomization for surveys are so hard to find. We see early discussions of urn models from time to time, but they are nearly always hypothetical: actual samplings are nearly always systematic when they are not purposive. We have not run across such devices as use of a random starting point

Kiaer is not well known among statisticians, but he has not been neglected in historical accounts. He might be better known if his contribution had led to a crisp formula, to a specific method, or to a book that delineated the field. He was at the front of a wave and probably speeded developments.

4 The American connection

Before swinging across the hiatus to 1925, we present a brief digression about Kiaer's two uses of American experiences and views as he argues for representative sampling.

Wilbur and vital statistics The first appears in Kiaer's 1899b, where he cites approvingly a then-recent article by Wilbur (1896–97) about USA vital statistics. Wilbur is troubled because in many parts of the United States birth, death and other vital statistics are not maintained. Those states that do have registration of vital statistics are not representative of the whole country; for example, the registration states (primarily in the northeast) have larger proportions of women, more elderly people, fewer Negroes, *etc.*

Wilbur proposes choosing 'small representative areas' (p. 195), perhaps counties, within the nonregistration states, and carrying out for those areas full registration of vital statistics. These would be 'typical localities' (p. 194), 'selected with reference to geographical, topographical, climatic, and other conditions, as to be fairly representative . . . of each State. Density of population, nationality, and prevailing occupations . . . should also enter the . . . selection. Death rates based upon the accurate registration in the several registration districts of a State would be a fair representative of the mortality of the State as a whole. [By population weighting] . . . a representative death rate for the United States could then be obtained . . .' (P. 196–97.) As usual in these proposals for purposive sampling, the goals are vigorously expressed, but the means of accomplishing those goals are missing.

In the ensuing published discussion, one argument against the suggestion is that of undue federal intervention, but most of the discussants were in favor. The next article in the journal, by William A. King, p. 209–220, goes into detail about how the 1900 Census could be applied to data from the 'representative non-registration areas' (p. 219) and gives a list of 'registration cities in nonregistration states' (p. 220) for the 1890 Census.

Wilbur's suggestion was proposed to Congress by Carroll D. Wright, to whose own writings we soon turn. We have not tried to learn what became of the representative district idea for American vital statistics. One thing is clear: the idea is not nearly as well thought out as Kiaer's representative method. In particular, Wilbur proposes *typical* districts, as against Kiaer's central

---

for systematic sampling in the early survey sampling literature, and the idea of several independent random starting points seems absent. (The first suggestion we know for several such random starting points was by Yates (1936–37).)

Yet the paucity of examples may not be so surprising, for one hardly ever finds physical randomization in survey samples *today*, and that statement holds even if one extends the scope of the sought after examples to include the use of tables of randomly generated numbers, tables that have been widely available now for decades. Even the United States Bureau of the Census, that exemplary organization, in its decennial census uses a systematic design for subsampling.

The arguments for systematic sampling today are presumably those of (1) cost savings, and (2) behavior resembling that of a probability sample. What might have been the arguments in the early days of the century? (We write 'might have been' for we have found no explicit discussion.) There would have been no practical problem, for gambling devices were readily available and the use of decision by lot for a variety of social purposes had had a long history. (See our footnote 3 and Fienberg (1971).) Random devices had been used for fortune telling and necromancy for millennia.

We speculate that the very fortune telling and gambling associations may have provided the barrier, together with the kindergarten-like character of picking slips out of hats, spinning pointers, etc. Participants at ISI meetings of the time strike us as men of considerable self-importance and dignity. The science of statistics was not firmly established, and associations with magic, with gambling, or with fun and games, would understandably be avoided. This is pure speculation, and we hope others will be able either to support it or to show the conjecture wrong.

idea of districts that provide *distributions* like that of the population. That ambiguity, in fact, pervades the entire history under discussion and is one central problem in dealing with the intuitive idea of representativeness.

Wright and labor statistics Now we turn to the second part of the American connection. R.G.D. Allen, in an encyclopedia biography of A.L. Bowley (1978) (p. 28) remarks that Kiaer was greatly supported during the 1901 Budapest meeting by the report of Carroll D. Wright 'on sampling experience in the U.S. Department of Labor'. Sure enough, on p. 70 Kiaer (1903) quotes two paragraphs of a letter to him from Wright. To paraphrase, Wright says

> The experience of the U.S. Department of Labour has continually strengthened my own views as to the value of representative statistics. [Nine annual reports] . . . are all emphatic evidences of the values of this method . . . . In fact, offices like [ours] . . . must use the representative method; and . . . on most of the topics with which they deal representative facts are quite sufficient, at least when we consider the vast cost of securing statements of aggregates.
>
> So, taking all things into consideration, . . the conclusions given in your St. Petersburg report and your more detailed study in 'Allgemeines statistisches Archiv' are eminently wise and sound.

How strong that support was is hard to tell. It may have had considerable rhetorical force – a fresh, strong wind from the New World of America – but unfortunately Wright, like Wilbur, expressed a superficial view of representative sampling in comparison with Kiaer's. So Kiaer, in drawing support from Wright, gained a testimonial that would not have stood up in vigorous debate. (So far as we know, no one raised that point in ISI discussions.)

The evidence that Wright's understanding was superficial comes from those writings specifically cited by Wright in the passage paraphrased above. We have read in the Annual Reports listed by Wright, and we have been awed by the enormous tabulations and summaries on such topics as industrial depressions, working women, wages, and hand and machine labor. Wright was clearly a vigorous, effective, articulate labor statistician, but he did not understand Kiaer's idea of representativeness.

Wright certainly used the rhetorical, dogmatic approach to representativeness. He simply asserts representativeness, for example,

> A canvass of a portion of the country, it was believed, would secure data fairly representative of the whole, and so nearly accurate as to provide a basis for estimating the figures for the whole country. Certain districts were, therefore, selected which were believed to be fairly representative of the whole country. (1898, p. 11.)

> These twenty-two cities, with the exception of San José, in California, must be considered as thoroughly representative, so far as locality is taken into account. (1899, p. 9.)

Wright's optimistic self-confidence is almost boundless, and typically unaccompanied by evidence. For example, in a paper outside the annual report series, he admits that errors may 'creep into an official report,' but 'such an occurrence . . . is exceedingly rare in the history of the Bureau' (1904) (p. 977). Thus most of Wright's usages of 'representative' come under our first meaning: general, if unjustified, acclaim for the data.

Wright does do a little comparison of marginals, but only in a fragmentary fashion. For example, in his Fourth Annual Report for 1888 (1889) (p. 62–63) he compares the average age for the women studied in Boston, 24 years and 5 months, with the corresponding average, 24 years and 9 months, for Boston women studied in 1883 by the Massachusetts Bureau of Statistics of Labor. This, he says, offers a 'satisfactory piece of evidence as to the representative character of the women interviewed'.

He also touches on the idea of coverage, again in a fragmentary way. For example, he says in his fifth Annual Report that it

> was impossible ... to cover in its entirety the railroad industry of the United States, but it was necessary to cover a sufficient number of roads representing all parts of the country and all conditions of railroad labor to make the investigation absolutely representative in character. (1890, p. 14.)

A concept of representativeness appears in Wright's work that was apparently new: a sample is taken to be representative if it includes a substantial fraction of the population. For example,

> ... the facts represent 56.29 per cent of the value of product, and must therefore be considered as absolutely representative of the conditions accompanying the production and supply of water, gas, and electricity by private and municipal establishments. (1890, p. 6.)

This concept by itself can hardly stand up to analysis. A large sample, nevertheless, can be representative in the sense of 'good enough for a particular purpose', as described in Section 2.9 of Kruskal and Mosteller (1979c). Thus a sample might simply show that some characteristic thought to be absent or rare in a population is in fact frequent, or, conversely, that a characteristic thought to be universal is in fact often missing.

So Wright's representativeness lacked two important features of Kiaer's: first, the method of sampling was hardly ever described or explained; and, second, Wright provides no systematic comparison of marginals with known information. On the other hand, Wright apparently understood the intuitive idea of an approximate miniature of the population, although we have not found that word used by him.

It is curious that Kiaer should rest part of his case upon a testimonial from an American statistician who did not fully understand the message from Norway. No doubt support is always welcome even if it rests on shaky foundations, and presumably Wright's extensive use of partial investigations, and his ebullient satisfaction with them, encouraged Kiaer.

Wright's career was a long and fascinating one; we recommend the biography by Leiby (1960). From it, for example (p. 115), we learn that Wright's ideas of representativeness in terms of large sample size and ample coverage, as well as his rhetorical orotundity, go back to the 1870's when he was Chief of the Massachusetts Bureau of Statistics of Labor. For example,

> We have aimed to make our investigations of such a degree of comprehensiveness that our deductions would bear the impress of true representative character, and seem founded upon a tangible basis. (Wright, 1875) (p. 201).

Indeed, our view of the American connection is much like that of Jensen (1926) in an article to which we soon shall turn,

> ... the official statistics in both the United States and Canada have made wide use of partial investigations as substitutes for complete statistics, but ... the methods used there ... [cannot] be termed 'representative' in the narrower sense which this expression ... ought to symbolise .... This applies to the numerous partial investigations ... made by the United States Department of Labor and by Canada's Dominion Bureau of Statistics, ... (Jensen, 1926b) (p. 427.)

We have two final remarks about Wright and his enthusiasm. The first is that Wright, Kiaer, and others often began their sampling procedures by choosing substantial units such as whole cities or regions. For example, in Wright's 1888 study of working women, he began with twenty-two cities that, with one exception, were 'thoroughly representative'. Now it may well be that for many of the variables they studied the average value of a characteristic varied little from one substantial unit to another. A similar point may apply to distributions. To the extent this is true, it may have mattered little what units (cities, *etc.*) were selected. Thus it may be that a partial investigation based on such chunks, even one poorly designed, can be accurate; the weakness lies in the difficulty of *appraising* accuracy.

Our second point refers to Wright's enthusiasm. The unstated issue is how well the investigations guided policy decisions and helped with the planning and the execution of programs or legislation. In 1980, it is fair to say that we have no general method of appraising the value of statistical information for these purposes, governmental or private. The lack of this methodology is a shameful gap that economists, statisticians, and information scientists ought to fill. As a governmental statistician, Wright had no guidance about the policy value of his statistics, accurate or not, and we are in much the same position today.

5 Between 1903 and 1925

As mentioned at the close of Section 3, the pages of the *Bulletin of the International Statistical Institute*, for reasons we do not understand, said little about representative sampling after the 1903 meeting until the 1925 meeting in Rome.

Petter Jakob Bjerve, now Director of the Central Bureau of Statistics of Norway, has helped us more nearly to understand how Kiaer was overloaded with growing statistical responsibilities and a limited staff. Bjerve adds, in a letter, that Kiaer's work in sampling was a 'one man show. When he retired, there seemed to be no one to carry on in this field . . . and little interest for sampling . . . until after the war.'

The literature of the period does, of course, show a substantial movement towards more carefully designed and documented sample surveys, together with some material about probabilistic sampling – but little that was new or that might serve to clarify the obscurities we have seen, obscurities that awaited Neyman's major 1934 paper for much of their lifting.

Karl Pearson The ideas already described certainly continued to be in the air, and we illustrate that with two quotations from Karl Pearson. The first reaches back a little to 1897 and uses the cousinly expression 'fair sample', which he explains in a footnote given here in square brackets.

> Death strikes all ages, sexes, and conditions, and the graveyard gives a tolerably fair sample of the general population. [A 'fair sample' will be a random sample, – one which has not been specially selected, and in which pathological variations will consequently be in a very small minority.] (1897) (p. 268.)

That quotation relates both to coverage and non selectivity. Coverage is the theme of the following quotation within a quotation:

> . . . the homes of all the children attending a certain school . . . were examined and the school was chosen because [says Miss Dendy's report] 'It has upon its rolls children from the poorest parts of the city, and yet it has also an admixture of the substantially comfortable and thoroughly respectable working class . . .. In the poorest part of a city of many centuries' growth there are also many 'old families' who continue to reside in the houses their fathers and grandfathers lived in . . . This gives the school a widely representative character . . ..' (Elderton and Pearson, 1910) (p. 5.)

Von Bortkiewicz We mention also a 1911 discussion by von Bortkiewicz (1912) whose earlier comments we have already described. Bortkiewicz distinguishes between private statistics and official statistics; the former tend to be selective and partial, the latter general and comprehensive. Among the dangers of private statistics is that of non representativeness.

> The sample on which the research is based and the population of interest may not be so related that one can take the sample as representative of the population. (Free translation from p. 169.)

In particular, says Bortkiewicz, the number of cases may be too small to permit eliminating chance effects by the law of large numbers, and one should use the formulas of mathematical statistics.

Representativeness may be examined by comparing sample marginals with those of a known population, in the manner of Kiaer, but this method may be misleading, because margins that agree do not imply joint distributions that agree. Bortkiewicz then treats the target population vs. the sampled population, a favorite theme of his: a sample of workers in a given factory, no matter how large, may not generalize to workers in other factories.

He also gives a striking discussion of non-representativeness arising from selection effects in retrospective observations: an extreme example is that of starting with a group of idiots and examining their parents and grandparents, naive interpretation might lead to the conclusion that the mental ability of the population has sharply declined.

Fredholm Dalenius (1957) calls our attention to a remarkably advanced sample survey in the field of forestry (1909–1911): Sweden wished to measure its forest reserves in Värmland, and the planning committee wanted the results to be within 10 per cent of the correct value. The committee's original plan was criticized and others proposed that error control be based upon the probability calculus. Then the great Swedish mathematician Ivar Fredholm of integral equation fame was asked for his judgment on the new proposal. Dalenius's (1957) (p. 46–67) translation of Fredholm's 1910 letter says,

> (1) I consider that the probability calculus can be applied to the estimation of the forest capital in the manner suggested, and that the laws valid for the calculus mentioned, which have been successfully applied in many other fields, can be applied to equally great advantage to the problem with which the Commission is faced.
>
> (2) It should be possible, using the material available within a given district, to draw conclusions regarding the probable error, to which the survey can be subjected, and in this manner determine whether the survey percentage has been sufficient or not for the need.
>
> (3) It should be possible, using the proposed preliminary survey of the County of Värmland, to decide whether a strip survey with the low percentage proposed can give results sufficiently accurate for the purpose required.

The final tree selection was systematic within strata which were strips of land and sizes of trees. The Värmland Survey apparently had important consequences for sample surveys in forestry, but its influence did not spread to other kinds of investigations.

Bowley A.L. Bowley dealt influentially with representativeness and social surveys in his pioneer work on income and other characteristics. His 1915 book, *Livelihood and Poverty*, with A.R. Burnett-Hurst, discusses representativeness (for example on p. 12) of his roughly one in 20 systematic sample of a purposively chosen sample of four English towns. Bowley's chapter VI deals with accuracy and is for the time a remarkable discussion. It touches on measurement error, ambiguity of concept, and sampling error. For sampling, it uses the mathematics of simple random sampling and proposes estimation by intervals around the average of plus and minus three times the probable error of sampling. This is roughly two standard errors, and hence provides rough 95 per cent confidence intervals. In fact, he almost – but not quite – comes on p. 180 to the concept of confidence intervals.

Then – and this is a considerable advance on Bortkiewicz and others – Bowley points out that in fact his systematic sampling gives him a *smaller* margin of error than his formula based on randomness.

> ... our samples were deliberately taken from houses distributed nearly uniformly through the towns, whereas the formula applies to samples taken quite at random, as if the houses had been numbered and numbers drawn from a bag. Our distribution cannot then contain a greater number of extreme cases: it must have a relation to the distribution of the rich ... and the destitute, so far as there are aggregated, similar families in similar streets. (P. 181–182.)

Bowley says that the systematic sampling reduces the 'margin of error' so that the intervals via randomness provide a 'very safe limit'.

One can hardly fault Bowley for neglecting in 1915 possible dangers from systematic patterns or autocorrelations in the population. After all, he was well beyond some present-day sample survey practitioners who merrily treat systematic samples as random without any concern.

A momentary jump ahead in time illustrates Bowley's sensitivity to sources of selection bias. John Hilton reported in 1928 the success of four sampling investigations of the industrial and insurance histories of claimants for unemployment benefits. During part of these studies, interviews were held with some claimants, following a plan to interview the first person to appear in the office with a file number within three of a previously sampled file. Bowley had earlier objected that this would create selection bias, but Hilton had thought that unlikely. In the 1928 paper, Hilton reported that the selection bias feared by Bowley had in fact occurred - claimants with allowed claims were required to appear and sign a register several times a week, but claimants with disallowed claims signed only once a week! Hence claimants with allowed claims were greatly over-represented.

Chuprov The publications of A.A. Chuprov (Tchuprow, Tschuprow, *etc.*) are worthy of special note, particularly his 1923 *Metron* two-part paper. Chuprov had an unusually clear idea of probabilistic structure and the growing importance of stochastic dependence among the observations, dependence of the kind that appears in sampling without replacement from a finite population as in a social survey. In particular, Chuprov was led to consider optimal allocation in stratified sampling in 1923, well before Neyman's independent and far more influential publication of the result in 1934.

Chuprov's work is discussed by Kaufmann (1913), especially on p. 367, and we cite two helpful obituary discussions by Isserlis, (1926) and Georgievski, (1928). Chuprov's influence appears to have been limited because of the formalities and abstractness of his style, and because of his geographical movements. See also our footnote 4.

Koren The American Statistical Association's Diamond Jubilee volume edited by Koren (1918) says very little about sampling though it describes statistical work in many countries. In the chapter on Norwegian statistics, Kiaer has five lines on sampling (p. 459). In the chapter on Russian statistics, A. Kaufmann mentions sampling (p. 525, 527, 528), but gives no details, except that he regards A. Pjeschechonow and Grohmann as the chief sponsors of the method, the latter 'has successfully attempted its theoretical defence'. (P. 528.)

Summary. Thus the major advances in understanding representative sampling during the period 1903 to 1925 may be ascribed to Chuprov and Bowley, but with the latter considerably more influential.

6 The 1925 Rome ISI meeting

The ISI 1925 Rome meeting (proceedings published in 1926) produced a great flowering of material on representativeness. We now summarize its most important points.

Jensen (1926a and 1926b) was leader and rapporteur for the discussion, which was framed in terms different from the older ISI debates. Time and generation had won the day for sampling, and the big questions now were its proper applications and the analysis of sample observations. The very title of the discussion was 'The application of the representative method,' and Jensen made the point (1926a) (p. 59) in these words,

> When the International Institute of Statistics discussed the matter twenty two years ago it was the question of the recognition of the [representative] method in principle that claimed most interest. Now it is otherwise . . . nowadays there is hardly one statistician, who in principle will contest the legitimacy of the representative method

Bowley was the other special leader of the discussion. We shall return to his theroretical development, but he is strong early on about matters of good statistical practice. In every sampling inquiry, he says,

> there must be . . . a complete absence of prejudice and a perfect willingness to accept the results, however unpalatable. . . . exactness of definition . . . must be obeyed with even more scrupulousness than in a complete investigation . . the universe from which the selection is made must be defined with the utmost rigour . . . the rules of selection, determined by the mathematical analysis, must be followed with perfect strictness. (Discussion of Jensen 1926a) (p. 62–63.)

Some might feel that Bowley does not fully follow his own precepts, but then who among us does? We shall return to Bowley. (In addition to Jensen and Bowley, the Commission which had been appointed in 1924 to study the representative method, had as members Corrado Gini, Lucien March, Verrijn Stuart, and Franz Žižek.)

The discussion led at last to a formal resolution (p. 68–69, 212–213, and 377–380) with carefully chosen wording to placate the last few holdouts. That resolution distinguishes two principal kinds of representative sampling: random sampling (with all elements of the population equally likely to be chosen) and purposive sampling of groups of population elements (much like cluster sampling, but apparently with large clusters, full enumeration of selected clusters, and judgment in making that selection).

The descriptions of random sampling are generally incomplete; they require that each unit have equal probability of appearing in the sample, but they do not speak to *joint* probability. The idea of looking at the distribution of the entire sample point still lies in the future. Except for the early start by March, the general idea of probability sampling also seems as yet unborn, although there is some discussion of stratified sampling, always with stratum sample sizes proportional to stratum population sizes (in order to maintain constant choice probability over units).

Two papers by Jensen (1926a and 1926b) provide extensive discussion; the first is a report of the ISI commission. Here he spells out the purposive sampling idea a bit more: the chosen groups should have average values approximately equal to population averages for those characteristics (control variables) already known for the population. This idea, in some ways natural, in other ways intrinsically vague, is intuitively connected with the motivation for covariance analysis, but it is still too ambiguous. Elsewhere Jensen requires that the groups be chosen so that the control variables have whole distributions like those in the populations – not just averages. This tension seems unresolved.

Jensen (1926a) clearly describes the non-selectiveness idea of representation, our meaning 2 (*e.g.*, p. 363), the typicalness idea, our meaning 4 (p. 369), and the rhetorically empty usage, our meaning 1 (p. 361).

In his second article (1926b), Jensen describes the representative method in practice at considerable length and gives a long bibliography that we have been unable to explore fully. This article includes a relatively explicit discussion of cluster sampling ("Random selection of groups"), although in execution the random selection seems instead generally to have been systematic. It also discusses purposive selection of large groupings, followed by random choice of units within the groupings. On p. 425–428, Jensen worries over the degree of representativeness of *his* sample of examples, a concern that we too have felt for ours.

March (1926) forms a link between the 1903 and the 1925 meetings. He contributes a short article on modes of sampling, with a brief bit of sampling theory.

At the end, Bowley (1926) gives a lengthy development that is the major theoretical discussion of the sequence. It recapitulates known material about sampling from finite populations, but breaks fresh ground in several ways. it almost reaches the confidence interval idea (p. 15!), goes further and almost reaches the multiple comparison idea – *i.e.*, the idea of a confidence region in a multinomial-like context via a chi-square statistic (p. 16!), and then provides a Bayesian treatment. Bowley attempts to make precise the purposive sampling idea in terms of correlations between control variates and the variate of primary interest, but he fails, in our reading, through lack of clarity in his description of sampling. As we shall soon see, Neyman, a few years later, also had a difficult time reading Bowley, and produced his own mathematization.[7]

Thus at the 1925 meeting the discussion is not whether to do sampling, but how to do it. Progress has appeared in stratified and even in cluster sampling. The idea of purposive sampling continues to live and is endorsed, although it remains vague.

7 Gini-Galvani and the Neyman watershed

The great Neyman article of 1934, 'On the two different aspects of the representative method . . .', requires attention in any historical treatment of the idea of representativeness. It begins with a description of the problem faced by Corrado Gini and Luigi Galvani of the Italian census, and so we too touch briefly on the Gini-Galvani ultimately failed attempt to use purposive sampling for a kind of sharp representativeness. Neyman himself has since retold the fascinating story at least once (1952) (p. 105–107; original edition 1938), so we may compress our reminder. The references are Gini (1928), Gini and Galvani (1929).

Gini and Galvani had to discard most of the original returns of an Italian census to make room for the next one, but they properly felt it essential to keep a sample of the older material to permit further studies. They wanted a representative sample in the then current sense (the date was 1926 or 1927), and so they decided to retain the raw data for 29 out of 214 administrative units (*circondari*) into which Italy was divided. Which 29? By trial and error, 29 *circondari* were found such that their average values for seven important characteristics (birthdate, death rate, altitude, *etc.*) were all close to the corresponding averages for the entire country. Yet when *other* characteristics were looked at, or even when aspects other than averages (*e.g.*, variability, association) were looked at for the primary seven characteristics, Gini and Galvani found wide deviations between the sample and the country as a whole. Gini concludes his 1928 article with a sharp criticism of sampling as such, almost a return to the older von Mayr approach. Neyman, on the other hand, argues that we should distrust purposive sampling, not sensible probability sampling.

The 1934 Neyman paper is chockablock full of so many important ideas and insights that we must restrain ourselves to deal only with representativeness. Thus we shall not discuss Neyman's treatment of the Bayesian approach, of confidence intervals (apparently Neyman's first treatment of that topic in English), and of the sample sizes that give optimal stratification. We shall also resist analysis of the discussion by Fisher, apparently the mild first of the increasingly bitter sequence of arguments between Fisher and Neyman.

[7] Yates (1946) summarizes the conclusions of the 1925 meeting, and outlines subsequent developments, with special attention to the use of analysis of variance and related ideas in the design and analysis of surveys. The paper includes a residual of the older control variable idea, now called balancing. It remains vague and of uncertain merit; perhaps for that reason it excited special discussion by M.G. Kendall, F.J. Anscombe, and M.H. Quenouille. In a 1979 letter, Yates tells us that 'I think in sample survey work, there is nothing to be said for attempting to get balanced samples. A much more satisfactory procedure is to adjust the results by means of regression. This procedure is similar to what I have termed stratification after selection.'

The Gini–Galvani example, together with Neyman's exposition of it, advertised the weakness of purposive sampling so that it became widely understood. The example turned most statisticians away from the idea of purposive sampling in sample surveys with a force that the intrinsic fuzziness per se of the concept never had. The notion of sampling of clumps is a different story, to which we now turn.

The second contribution of Neyman's paper to our theme is its insistence on exact specification of the sampling procedure. For purposive sampling as treated by Bowley and Gini, the procedures are vague and it is not clear whether or how probability enters. As Neyman says (p. 573), 'I could not exactly follow the method proposed by Professor Bowley'. In order to obtain precision, Neyman envisages the universe of districts (*i.e.*, mutually exclusive groups of units, whose union makes up the population of individuals, households, *etc.*) as divided into strata such that within each stratum (1) the value of the control variable(s) is (are) approximately constant, and (2) the number of units is approximately constant. Then, within each of those strata, a preassigned number of districts is chosen at random. The preassignment, regarded as fixed and decided in advance, is such as to give an average value for each control variable approximately equal to the corresponding population average.

By this device, the problem is stated as stratified sampling with one or more covariates. With a linear regression relationship assumed, one can then proceed to analysis; Neyman gives a detailed example.

If the districts are large, as in the Italian case, then most of the strata contain zero districts or one. This implies a large sampling bias or a large sampling variance or both, depending on just how the estimation is approached, and it points to the central problem for Gini and Galvani even after Neyman's plan is created.

To summarize, although Neyman gave an impressive solution to the problem of the purposive sampling of clumps, what was probably more important was his and Gini and Galvani's exposition of the failure of one method of purposive sampling in a practical large problem. Purposive sampling lost its appeal.

8 The 1935–1939 period

Neyman, Fisher Our roughly chronological exposition ends just before 1940. It seems reasonable to make the cut there, because by then the Neyman watershed paper had been well circulated and absorbed; in addition, World War II introduced both enormous disruptions and stimuli to statistical thinking

In an important sense, Neyman's (1934) paper brought us into modern thought in many contexts, including those related to surveys where the representativeness idea arises so often. We cite also his 1937 Department of Agriculture lectures, published in 1938 and later, in considerably revised form, in 1952. Those lectures derived from initiative by W. Edwards Deming, then head of the statistical program at the Department of Agriculture's Graduate School ... and a tireless advocate of high standards in sampling.

As far as we know, Fisher did not use the term 'representative sample', although the more important senses of the term are of course central in much of Fisher's work. For example, protection against unwanted selection in sampling is a major motivation for randomization in design (although not for Fisher the principal motivation). Indeed the selection idea is central in Fisher's explication of the meaning of probability, and we see a warning against selection in the following statement from Fisher (1936):

> ... the student of living measurements can choose his material and be sure of getting enough of it. He is not dependent upon such an accident as that the cemetery of a community, perhaps not well representative of the racial type of its neighbours, has happened to be excavated with unusual thoroughness. (P. 63.)

The quotation hints at the need for adequate sample size, and on page 60 of the same paper a remark suggests a close linkage between representativeness and having a random sample.

Marginal comparisons Empirical probing of representativeness by comparison of sample marginal distributions with known population marginals was and is frequently found. A 1932 example by Margaret Hogg appears in Appendix I, 'Evidence of representativeness of sample population studied,' of her book on work shortage. She compares her sample with the 1930 Census as to

sex by age, total population
sex by age, gainful worker population
sex by marital status
sex by kind of industry for gainful workers
sex by nativity, total population
sex by nativity, gainful workers
wards, for population 15 and over

Although such comparisons are desirable, they have the intrinsic logical difficulty that even a perfect marginal match between sample and population implies only a little about joint relationships. Although marginals never match perfectly, Kiaer would be pleased with the wide use of matching exemplified by Hogg, including bivariate marginals. Hogg uses estimated standard deviations and a kind of significance testing procedure that we do not fully understand. Earlier in the book (p. 13–19), Hogg compares random sampling with judgment sampling, which she calls 'judicial selection', and strongly advocates random sampling.

Another author of this period, Clyde V. Kiser, took a similar approach in a 1934 article. Representativeness of a sample of households for studying morbidity rates was judged by comparisons with the Census for several characteristics, primarily size of household. The sample had too few smaller households and too many larger ones, presumably for two reasons: first, secondary households (*e.g.*, a married daughter with her family in the same house as her parents) were not recorded; and second, failure to revisit not-at-home households undoubtedly acted selectively to exclude more small families than large ones. A further enumeration tried to fill in these exclusions, and the resulting household size distribution was close to that of the Census.

The customary lingering questions remain. Kiser remarks that the study '... had been primarily concerned with securing a sample representative of the area in regard to prevalence of sickness rather than size of household' [p. 251]. But then how do we know that improvement on the latter characteristic improves the former? Only in general terms because it must be helpful to remove two sources of selection or bias. Might there be others? One can never be certain, and Kiser does not describe his sampling procedure.

Confusion in the 30's is further exemplified by the statistical articles in the *Encyclopaedia of the Social Sciences*. The article on statistical practice, by Robert M. Woodbury, shows (1934) (p. 364) a mixture of vagueness and dogmatism about representativeness. Oskar N. Anderson's article on statistical method, although still with a touch of circularity, is (1934), (p. 369–370) far clearer. One pair of papers from that period shows the explicit influence of Neyman's (1934) article. The two papers are by Stouffer (1934 and 1935); it seems clear from internal evidence that Stouffer had read Neyman's paper between writing his own two. See especially p. 508 of the 1935 paper.

Selection bias Two important papers in the mid-30's form a link between the biological and experimental developments in which R.A. Fisher played so central a role and the survey-oriented thinking that we have described at length. The two papers are by Yates (1934/35) and by Cochran and Watson (1936). Both empirically explore selection bias introduced by choice of plants for measurement by expert eye. Representativeness is much in the authors' minds; a key early paragraph by Yates says

> The ideal . . . in sampling is to make the sample as representative as possible, so that measurements or observations on it can be taken as virtually equivalent to similar measurements on the whole population. . . . this ideal . . . in the mind of the sampler . . . naturally influences his selection if he has any freedom of choice. Most samplers when selecting a representative sample will deliberately reject abnormal material, or if they feel that the sample should be representative of the abnormal as well as the normal will deliberately balance up the different categories of abnormality. (P. 202.)

One sees in that short passage the ideas of a population miniature, the typical case, and coverage, our meanings 3, 4, and 5, respectively. Then Yates comes to his central theme of selection bias,

> Unfortunately the sampler's claims to be able to select 'a representative sample' by personal judgment are largely unfounded, and his selection is in fact subject to all sorts of biases, psychological and physical. (P. 202.)

He then preaches the good doctrines of objective sampling methods, replication, and randomization. Yates goes on to much further instructive advice, including

> Perhaps . . . the most glaring examples of bias in human sampling are provided by postal enquiries or questionnaires. Since only those . . . interested in the problem . . . are likely . . . to reply . . . such enquiries will be entirely unrepresentative unless very special efforts are made to ensure a complete set of answers. (P. 213.)

We note in passing that Yates, writing in the then-new glow of enthusiasm for Fisher's ideas about design, also comments favorably about Neyman's fundamental 1934 paper and quotes from it.

Most of Yates's article deals with three specific agricultural examples in some detail. The first analyzes an experiment on height of wheat shoots (and ears) in which random sampling was used *plus* judgment sampling to bring the sample to desired size. The measurements were fortunately labeled so that one could compare the randomly and judgmentally chosen plants. Differences were striking. When the shoots were low, the judgmentally selected ones tended to be too tall; when the shoots were in ear, the judgmentally selected ears tended to be too near the average, avoiding both extremes but not symmetrically (p. 208).

The article by Cochran and Watson (1936) describes an experiment on selection bias motivated by the Yates study. Twelve trained observers – participants in a 1935 agricultural conference – were asked to choose judgmentally samples of wheat shoots. All observers chose samples of average height greater than the true mean height. The amount of selection bias varied considerably over observers and over sampling unit plots; in addition observers tended to avoid exceptionally tall and short shoots. A new discovery (although a familiar staple in the study of the failure of the personal equation among astronomers, Boring (1929)) was that individual observers were not consistent through the experiment: the bias was greater when true height was greater, so that immediate estimated differences between varieties would be distorted.

A third paper in this line, by Yates (1936–37), contains a number of valuable contributions, including discussion of experimental sampling by Cochran in which observers selected from 1,200 stones on a table samples that 'represent as nearly as possible the size distribution of the whole collection'. (P. 5.) As might be expected, the observers tended to pick stones on the large side and to show unduly small intra-sample variabilities.

Quota sampling In our third paper (Kruskal and Mosteller, 1979c) (p. 249–250 and 258), we discussed quota sampling, a kind of purposive sampling that attempts to mimic the population in particular respects. Like all purposive sampling, quota procedures suffer from potential bias because similarity to the population in some respects by no means implies similarity in others. In addition, it is impossible to give sampling errors for estimators based on quota sampling except by the sometimes awkward method of replicating the entire sampling procedure. On the other hand, careful quota sampling is often used to good effect, and it is said by some to be less costly at first than probability sampling. Clear descriptions of quota sampling are given by Stock (1947) (p. 139–141), by McCarthy (1949) (p. 87–91), and by others. Terminology varies; for example, Stock uses the expression 'controlled sampling'.

Ironically, while Neyman was successfully attacking purposive sampling in professional statistical publications, there began to flourish in public opinion and marketing circles of the United States an independent growth of quota sampling. The basic idea, after all, is so appealing: establish a sample that is an 'America in microcosm', as Roper (1940) (p. 326) said, a microcosm in terms of region, size of place, sex, age, occupation, economic level, etc., and then ask the members of that microcosm their views; surely one will obtain a miniature of the views of everyone. Gallup (1948) (p. 22–23) put it thus: 'If a sample is accurately selected, it represents a near replica of the entire population. It is a miniature electorate with the same proportion of farmers, doctors, lawyers, Catholics, Protestants, old people, young people, businessmen, laborers, and so on, as is to be found in the entire population. [Such] samples . . . reflect opinions of the entire nation, with the small margin of error described earlier.' These quota methods had substantial successes in election forecasts until 1948, when they failed badly in the Dewey-Truman presidential election. Yet quota sampling alone did not seem to be the key assignable cause; there were many causes. See Mosteller *et al.* (1949). In view of the pre-1948 successes, it may not be surprising that the quota sampling literature contains so little reference to the kind of analysis provided by Neyman in the mid 30's and disseminated widely.

Quota sampling in public opinion polling and marketing went on for many years, and still continues, without much interaction with the work of academic or otherwise organized statistics. A description is given in Parten (1950) (Chapter 1); see also Duncan and Shelton (1978) Chapter 2, especially p. 36. Of course there has been criticism by statisticians, for example, see Hauser and Hansen (1944), but the issues have seldom been squarely joined. An excellent balanced review is given by Stephan and McCarthy (1958) (p. 37–38 and chapters 8, 10, 11, and 12).

Final remarks Although it extends beyond our time window, we wish to cite the pioneering work on sampling at the U.S. Bureau of the Census by Morris H. Hansen and others in the late 30's and thereafter. This superb work on both the theory and practice of large-scale, complex probability sampling is described in all current sampling textbooks; its institutional history is outlined by Duncan and Shelton (1978) (Chapter 2). See also Hansen and Madow (1976).

We close this historical sketch with a description of a 1939 paper by George W. Snedecor. It begins with the statement, 'The history of sampling is the history of statistics', and it ends with an argument for replacement of censuses by sample surveys! Snedecor gives an excellent review of the history that we have described at greater length, he complains (p. 853) about the purely rhetorical use of 'representative sample', our meaning 1, and he naturally discusses Fisher's work. In particular he says that

> [Fisher's] epoch making discovery of the *z* distribution together with the associated . . . analysis of variance furnish a method . . . for treating . . . many stratified random samplings. This method, quickly adopted in [biology], . . . has received meager attention by investigators of social phenomena. Since the stratified random sample is recognized by mathematical statisticians as the only practical device now available for securing a representative sample from human populations, the use of analysis of variance . . . [should] increase rapidly in social researches. (P. 852.)

How times change!

Acknowledgements

It is our pleasure to acknowledge the enthusiastic support we have received from many colleagues around the world. We have been astonished at the frequency and promptness of responses. Some have supplied us with reference materials and oral history and guided us to other scholars: Petter Jakob Bjerve, William G. Cochran, Jean Converse, W. Edwards Deming, Otis Dudley Duncan, David Durand, Lester Frankel, Martin Frankel, J. Arthur Greenwood, Morris Hansen, William Jaffé, Zoltan Kenessey, Leslie Kish, William G. Madow, Arnaldo Momigliano, Lincoln Moses, Jerzy Neyman, Leon Pritzker, Paul B. Sheatsley, O.B. Sheynin, Susan Shott, Frank Yates, Sandy L. Zabell.

Many have generously criticized versions of the manuscript: Keith Baker, Miriam Gasko Green, David Hoaglin, Patricia Hume, Joseph B. Kruskal, Martin D. Kruskal, Thomas Louis, John W. Pratt, Harry V. Roberts, Nancy Romanowicz, I.R. Savage, Conrad Taeuber, Judith Tanur, Ronald Thisted.

We owe special thanks for very extensive suggestions, not all of which we could handle, by Robert Boruch, Tore Dalenius, David L. Sills, Stephen M. Stigler.

For extensive and rapid translations from German, we were helped by Thomas Huber and Tanja Kazan. Christine Waternaux kindly contributed the French summary. The pictorial mnemonics for the meanings of representative sampling were prepared by Martha Carn. Finally for work on all aspects of the effort, we owe a profound debt to Cleo Youtz.

Preparation of this paper was aided by National Science Foundation grants NSF SOC 72-05228 A04 to the University of Chicago and NSF SOC 75-15702 A03 to Harvard University. One of us (WK) held a fellowship from the John Simon Guggenheim Foundation during part of the preparation of this paper, and that aid is gratefully acknowledged.

References

Allen, R.G.D. (1968). Arthur Lyon Bowley, *International Encyclopedia of the Social Sciences*, edited by David L. Sills. Macmillan and Free Press, New York. Also p. 27–30 of *International Encyclopedia of Statistics*. (1978), edited by William H. Kruskal and Judith M. Tanur. Macmillan and Free Press, New York. 2, 134–137.

Anderson, Oskar N. (1934). Statistical Method. *Encyclopaedia of the Social Sciences*, edited by Edwin R.A. Seligman and Alvin Johnson. Macmillan, New York. 14, 366–371.

Boring, Edwin G. (1929; corrected second printing in 1931). *A History of Experimental Psychology*. Century, New York. 1950, second edition. Appleton, New York.

Bortkiewicz, L. von (1912). Discussion in 'Probleme der Arbeiterpsychologie unter besonderer Rücksichtnahme auf Methode und Ergebnisse der Vereinserhebungen'. *Verhandlungen des Vereins für Sozialpolitik, Schriften des Vereins für Sozialpolitik*. 138, 115–204. Report of a 1911 meeting. Bortkiewicz's discussion is on p. 168–174 and 197–198.

Bowley, A.L. and Burnett-Hurst, A.R. (1915). *Livelihood and Poverty*. G. Bell, London.

Bowley, A.L. (1926). Measurement of the precision attained in sampling. *Bulletin of the International Statistical Institute*, 22, Supplement to Liv. 1, [6]–[62]. Preceded by summaries in French and English.

Chang, Wei-Ching (1976). Sampling theories and sampling practice. In Owen, D.B., *On the History of Statistics and Probability*. Dekker, New York, p. 299–315.

Cochran, William G. (1978). Laplace's ratio estimator. *Contributions to Survey Sampling and Applied Statistics: Papers in Honor of H.O. Hartley*, edited by H.A. David. Academic Press, New York, p. 3–10.

Cochran, W.G. and Watson, D.J. (1936). An experiment on observer's bias in the selection of shoot-heights. *Empire Journal of Experimental Agriculture*. 4, 69–76.

Dalenius, T. (1953). Representativitet. *Statistisk Tidskrift*. Statistiska centralbyrån. Stockholm. Swedish with English summary, 274–284.

Dalenius, Tore (1957). *Sampling in Sweden. Contributions to the Methods and Theories of Sample Survey Practice*. Almqvist and Wiksell. Stockholm.

Dewey, Frederic P. (1913). *The Sampling of Gold Bullion*. Bureau of the Mint, Treasury Department, Document No. 2672. Washington: Government Printing Office.

Duncan, Joseph W. and Shelton, William C. (1978). *Revolution in U.S. Government Statistics, 1926–1976*. Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, Washington.

Edin, K.A. (1912). De mindre bemedlades bostadsförhållanden (första bostadsbeskrifningen) i Göteborg. Statistisk undersökning utförd år 1911 på uppdrag af Kommittén för stadens kommunalstatistik. Göteborg.

Ehrenberg (1927). Losung. Kroll, Wilhelm and Wissowa, Georg, *Pauly's Real-Encyclopädie der Classischen Altertumswissenschaft*, neue Bearbeitung. Metzler, Stuttgart, 13, half-volume 26, p. 1451–1504.

Elderton, Ethel M., with the assistance of Karl Pearson (1910). *A First Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring*. Eugenics Laboratory, Memoirs. University of London, 10.

Fiebiger (1901). Decimatio. Wissowa, Georg, *Pauly's Real-Encyclopädie der Classischen Altertumswissenschaft*, neue Bearbeitung. Metzler, Stuttgart. 14, half-volume 8, p. 2272.

Fienberg, Stephen E. (1971). Randomization and social affairs: the 1970 draft lottery. *Science*. 171, 255–261.

Fisher, R.A. (1936). 'The coefficient of racial likeness' and the future of craniometry. *Journal of the Royal Anthropological Institute*. 66, 57–63. In Fisher's *Collected Papers*. III, No. 141, 484–490.

Gallup, George (1948). *A Guide to Public Opinion Polls*. Princeton University Press.

Georgievski, P. (1928). Tchouproff Alexandre. *Bulletin of the International Statistical Institute*, 23, Liv. 1, 345–349.

Gini, Corrado (1928). Une application de la méthode représentative aux matériaux du dernier recensement de la population italienne (1er décembre 1921). *Bulletin of the International Statistical Institute*, **23**, Liv. 2, 198–215.

Gini, C. and Galvani, L. (1929). Di una applicazione del metodo rappresentativo all'ultimo censimento italiano della popolazione (1° dicembre 1921). *Annali di Statistica*, Series 6, **4**, 1–107.

Hansen, Morris H. and Madow, William G. (1976). Some important events in the historical development of sample surveys. Owen, D.B., *On the History of Statistics and Probability*. Dekker, New York, p. 75–102.

Hauser, Philip M. and Hansen, Morris H. (1944). On sampling in market surveys. *Journal of Marketing*, **9**, 21–32.

Heyde, C.C. and Seneta, E. (1977). *I.J. Bienaymé: Statistical Theory Anticipated*. Springer, New York.

Hill, Austin Bradford (1966). *Principles of Medical Statistics*. Eighth edition. Oxford, New York.

Hilton, John (1928). Some further enquiries by sample. *Journal of the Royal Statistical Society*, **91**, 519–530. Discussion 531–540.

Hogg, Margaret H. (1932). *The Incidence of Work Shortage*. Russell Sage Foundation, New York.

Huber, Michel (1933). Lucien March, 1859–1933. *Journal de la Société de Statistique de Paris*, **74**, 269–280.

[Isserlis, Leon; actually given as L.I.] (1926). Alexander Alexandrovich Tschuprow. *Journal of the Royal Statistical Society*, **89**, 619–622.

Jaffé, William (1965). Biography and economic analysis. *Western Economic Journal*, **3**, 223–232.

Jensen, Adolph (1926a). Report on the representative method in statistics. *Bulletin of the International Statistical Institute*, **22**, Liv. 1, 359–380. The meeting was in Rome in 1925. Extensive discussion appears on 58–69, 185–186, and 212–213.

Jensen, Adolph (1926b). The representative method in practice. *Bulletin of the International Statistical Institute*, **22**, Liv. 1, 381–439.

Kaufmann, Al. (1913). *Theorie und Methoden der Statistik*. Mohr, Tübingen.

Kenessey, Zoltán (1963). To the history of sampling in official statistics. United Nations Regional Seminar on Sampling Methods in Current Statistics, 16–27 September 1963, Budapest. 25 pages. Published in German, Zur geschichtlichen Entwicklung des Stichprobenverfahrens in der amtlichen Statistik. *Allgemeines Statistisches Archiv*, **48** (1964), 312–330.

Kiaer, A.N. (1895–6). Observations et expériences concernant des dénombrements représentatifs. Discussion appears in Liv. 1, XCIII–XCVII. The meeting was in Berne in 1895. *Bulletin of the International Statistical Institute*, **9**, Liv. 2, 176–183.

Kiaer, A.N. (1897). Den repraesentative Undersøgelsesmethode. *Christiania Videnskabsselskabets Skrifter*. II Historiskfilosofiske Klasse, No. 4, 24 pages. Translated by Svein Bretvik into English, as 'The representative method of statistical surveys'. *Papers from the Norwegian Academy of Science and Letters*. II. The Historical, Philosophical Section. The original and the translation, together with a bibliography of Kiaer and a preface, were published in 1976 by the Norwegian Central Bureau of Statistics, No. 27 in its series *Samfunnsøkonomiske Studier*.

Kiaer, A.N. (1899a). Sur les méthodes représentatives ou typologiques appliquées à la statistique. *Bulletin of the International Statistical Institute*, **11**, Liv. 1, 180–185. The meeting was in St. Petersburg in 1897.

Kiaer, A.N. (1899b). Die repräsentative Untersuchungsmethode. *Allgemeines Statistisches Archiv*, **5**, 1–22. The author's name is spelled Kiär.

Kiaer, A.N. (1903). Sur les méthodes représentatives ou typologiques. *Bulletin of the International Statistical Institute*, **13**, Liv. 1, 66–70. Discussion 70–78. The meeting was in Budapest in 1901.

Kiaer, A.N. (1905). Untitled speech on the representative method. *Bulletin of the International Statistical Institute*, **14**, Liv. 1, 119–126. Discussion 126–134. The meeting was in Berlin in 1903.

King, William A. (1896/97). The vital statistics of the census. *Quarterly Publications* [later *Journal*] *of the American Statistical Association*, **5**, 209–220.

Kiser, Clyde V. (1934). Pitfalls in sampling for population study. *Journal of the American Statistical Association*, **29**, 250–256.

Koren, John, editor (1918). *The History of Statistics: Their Development and Progress in Many Countries*. Memoirs to commemorate the Seventy Fifth Anniversary of The American Statistical Association. Macmillan, New York.

Kruskal, William H. (1978a). Significance, tests of. *International Encyclopedia of Statistics*, edited by William H. Kruskal and Judith M. Tanur. Macmillan and Free Press, New York, p. 944–958.

Kruskal, William H. (1978b). Statistics: the field. *International Encyclopedia of Statistics*, edited by William H. Kruskal and Judith M. Tanur. Macmillan and Free Press, New York, p. 1071–1093.

Kruskal, William and Mosteller, Frederick (1979a). Representative sampling, I: Non-scientific literature. *International Statistical Review*, **47**, 13–24.

Kruskal, William and Mosteller, Frederick (1979b). Representative sampling, II: Scientific literature, excluding statistics. *International Statistical Review*, **47**, 111–127.

Kruskal, William and Mosteller, Frederick (1979c). Representative sampling, III: The current statistical literature. *International Statistical Review*, **47**, 245–265.

Lazarsfeld, Paul F. (1961). Notes on the history of quantification in sociology: trends, sources and problems. *Isis*, **52**, part 2, 277–333. Also pages 147–203 in *Quantification*, edited by Harry Woolf, Bobbs-Merrill, Indianapolis, Ind., 1961. Also, again, pages 213–269 of Volume II of *Studies in the History of Statistics and Probability*, edited by Maurice Kendall and R.L. Plackett. Griffin, London, 1977.

Lécuyer, Bernard and Oberschall, Anthony R. (1978). Social research: the early history of. *International Encyclopedia of Statistics*, edited by William H. Kruskal and Judith M. Tanur. Macmillan and Free Press, New York, p. 1013–1031.

Leiby, James (1960). *Carroll Wright and Labor Reform: The Origin of Labor Statistics*. Harvard, Cambridge, Massachussetts.

Lenin, V.I. (1956; first published 1899). *The Development of Capitalism in Russia*. Progress, Moscow.

Lichtenstein, Murray and Rabinowitz, Louis Isaac (1971–72). Lots. *Encyclopaedia Judaica*. Macmillan, New York, p. 510–514.

McCarthy, Philip J. (1949). The cross-sections used in predicting the 1948 elections. P. 81–118 in Mosteller, Frederick; Hyman, Herbert; McCarthy, Philip J.; Marks, Eli S.; and Truman, David B., *The Pre-election Polls of 1948*. Bulletin 60. Social Science Research Council, New York.

March, Lucien (1908). Remarques sur la terminologie en statistique. *Journal de la Société de Statistique de Paris*, **49**, 290–296.

March, Lucien (1909). De l'application de procédés mathématiques à la comparaison des statistiques. *Bulletin of the International Statistical Institute*, **18**, Liv. 1, 254–264.

March, Lucien (1912). Le traitment statistique des mesures mentales. (Review of *The Essentials of Mental Measurement* by W. Brown.) *Bulletin de l'Institut Général Psychologique*, **12** (No. 4), 227–275.

March, Lucien (1926). Observations sur la méthode représentative et sur le projet de rapport relatif à cette méthode. *Bulletin of the International Statistical Institute*, **22**, Liv. 1, 444–451.

Neyman, Jerzy (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, **97**, 558–606. Discussion 607–625.

Neyman, Jerzy (1952). *Lectures and Conferences on Mathematical Statistics and Probability*. Graduate School, U.S. Dept. of Agriculture, Washington, D.C. Second edition, revised and enlarged.

Nixon, J.W. (1960). *A History of the International Statistical Institute, 1885–1960*. International Statistical Institute, The Hague.

Parten, Mildred (1950). *Surveys, Polls, and Samples: Practical Procedures*. Harper, New York.

Pearson, Karl (1897). *The Chances of Death and Other Studies in Evolution*. Edward Arnold, London. Volume I. The quotation in text appears in the essay 'Variation in man and woman'.

Pearson, Karl (1928). On a method of ascertaining limits to the actual number of marked members in a population of given size from a sample. *Biometrika*, **20A**, 149–174.

Pearson, Karl (1978). *History of Statistics in the 17th and 18th Centuries*. Lectures, 1921–1933, edited by E.S. Pearson. Griffin, London.

Rabinovitch, Nachum L. (1973). *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. University of Toronto Press.

Rickard, T.A. (1904). *The Sampling and Estimation of Ore in a Mine*. The Engineering and Mining Journal, New York and London.

Roper, Elmo (1940). Sampling public opinion. *Journal of the American Statistical Association*, **35**, 325–334.

Seng, You Poh (1951). Historical survey of the development of sampling theories and practice. *Journal of the Royal Statistical Society*, Series A, **114**, 214–231. Reprinted in volume **II** of *Studies in the History of Statistics and Probability*, edited by Maurice Kendall and R.L. Plackett. Griffin, London, 1977.

Sheynin, O.B. (1976). P.S. Laplace's work on probability. *Archive for History of Exact Sciences*, **16**, 137–187.

Snedecor, George W. (1939). Design of sampling experiments in the social sciences. *Journal of Farm Economics*, **21**, 846–855.

Stephan, Frederick F. (1948). History of the uses of modern sampling procedures. *Journal of the American Statistical Association*, **43**, 12–39.

Stephan, Frederick F. and McCarthy, Philip J. (1958). *Sampling Opinions: An Analysis of Survey Procedure*. John Wiley, New York.

Stigler, Stephen M. (1977). Eight centuries of sampling inspection: the trial of the Pyx. *Journal of the American Statistical Association*, **72**, 493–500.

Stock, J. Stevens (1947). Some general principles of sampling. *Gauging Public Opinion*, edited by Hadley Cantril. Princeton University Press, chapter X, 127–142.

Stouffer, Samuel A. (1934). Sociology and sampling. *The Fields and Methods of Sociology*, edited by L.L. Bernard. Farrar and Rinehart, New York, 476–488.

Stouffer, Samuel A. (1935). Statistical induction in rural social research. *Social Forces*, **13**, 505–515.

Tschuprow, Al. A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, **2** (3), 461–493 and 646–683.

Westergaard, Harald (1932). *Contributions to the History of Statistics*. King, London.

Wilbur, Cressy L. (1896–97). Vital statistics for the twelfth United States Census. *Quarterly Publications* (later *Journal*) *of the American Statistical Association*, **5**, 188–200; discussion 200–208.

Woodbury, Robert M. (1934). Statistical practice. *Encyclopaedia of the Social Sciences*, edited by Edwin R.A. Seligman and Alvin Johnson. Macmillan, New York, **14**, 360–366.

Wright, Carroll D. (1875). *Sixth Annual Report of the Bureau of Statistics of Labor, March 1875*. Wright & Potter, Boston, Massachusetts.

Wright, Carroll D. (1889). *Working Women in Large Cities*. Fourth Annual Report of the Commissioner of Labor/1888. Government Printing Office, Washington.

Wright, Carroll D. (1890). *Railroad Labor*. Fifth Annual Report of the Commissioner of Labor/1889. Government Printing Office, Washington.

Wright, Carroll D. (1898). *Economic Aspects of the Liquor Problem*. Twelfth Annual Report of the Commissioner of Labor/1897. Government Printing Office, Washington.

Wright, Carroll D. (1900). *Water, Gas, and Electric Light Plants Under Private and Municipal Ownership.* Fourteenth Annual Report of the Commissioner of Labor 1899. Government Printing Office, Washington.

Wright, Carroll D. (1904). The Working of the United States Bureau of Labor. *Bulletin of the Bureau of Labor.* Based on an article in *Cosmopolitan Magazine* for June 1892. Department of Commerce and Labor. 9, 973–989.

Yates, F. (1934/35) Some examples of biased sampling. *Annals of Eugenics,* **6**, 202–213.

Yates, F. (1936–37). Applications of the sampling technique to crop estimation and forecasting. *Transactions of the Manchester Statistical Society.* Articles separately paginated; this has 26 pages.

Yates, F. (1946). A review of recent statistical developments in sampling and sampling surveys. *Journal of the Royal Statistical Society,* **109**, 12–30. Discussion 31–42.

Zabell, Sandy (1976). Review of Rabinovitch (1973). *Journal of the American Statistical Association,* **71**, 996–998.

Zahn, Friedrich (1934). *50 Années de l'Institut International de Statistique.* International Statistical Institute, Munich.

Žarković, S.S. (1956). Note on the history of sampling methods in Russia. *Journal of the Royal Statistical Society,* Series A, **119**, 336–338.

Zarkovich, S.S. (1962). A supplement to 'Note on the history of sampling methods in Russia'. *Journal of the Royal Statistical Society, Series A,* **125**, 580–582. These two articles have been reprinted in volume II of *Studies in the History of Statistics and Probability,* edited by Maurice Kendall and R.L. Plackett. Griffin, London, 1977.

Résumé

Nous suivons la trace des idées d'échantillonnage représentatif dans l'histoire de la statistique, commençant par la défense de A. N. Kiaer en 1895 en faveur de ce qu'il appelle la méthode representative et finissant juste avant la seconde guerre mondiale. Un sujet dominant est celui de la lente apparition d'idées probabilistiques éclairées dans le domaine des enquêtes statistiques, une évolution qui est encore loin d'être terminée. D'autres sujets abordés ici comprennent (1) l'ancien débat entre les statisticiens soutenant la méthode d'échantillonage et ceux qui insistaient sur une description complète, (2) les premières discussions – en général superficielles – de la représentativité en Amérique et leur utilisation par Kiaer comme arguments favourables, et (3) l'expérience de Gini et Galvani, et les rôles de Bowley, Neyman, et d'autres statisticiens.