Sci Eng Ethics (2013) 19:1057–1070
DOI 10.1007/s11948-012-9415-2

ORIGINAL PAPER

Karl Pearson and Eugenics: Personal Opinions and Scientific Rigor

Darcie A. P. Delzell · Cathy D. Poliak

Received: 23 August 2012 / Accepted: 4 November 2012 / Published online: 21 November 2012

Abstract The influence of personal opinions and biases on scientific conclusions is a threat to the advancement of knowledge. Expertise and experience does not render one immune to this temptation. In this work, one of the founding fathers of statistics, Karl Pearson, is used as an illustration of how even the most talented among us can produce misleading results when inferences are made without caution or reference to potential bias and other analysis limitations. A study performed by Pearson on British Jewish schoolchildren is examined in light of ethical and professional statistical practice. The methodology used and inferences made by Pearson and his coauthor are sometimes questionable and offer insight into how Pearson's support of eugenics and his own British nationalism could have potentially influenced his often careless and far-fetched inferences. A short background into Pearson's work and beliefs is provided, along with an in-depth examination of the authors' overall experimental design and statistical practices. In addition, portions of the study regarding intelligence and tuberculosis are discussed in more detail, along with historical reactions to their work.

Keywords Eugenics · Immigration · Jewish · Prejudice · Tuberculosis · Intelligence

D. A. P. Delzell (✉)
Department of Mathematics and Computer Science, Wheaton College, 501 College, Ave., Wheaton, IL 60187, USA
e-mail: darcie.delzell@wheaton.edu

C. D. Poliak
Sheldon B Lubar School of Business, University of Wisconsin Milwaukee, 2200, Kenwood Blvd., P.O. Box 413, Milwaukee, WI 53201-0413, USA

Introduction

In April of 2011, the editor of the *Annals of Human Genetics* made the archives of its first 29 years' worth of content available freely online, when the journal was known as the *Annals of Eugenics*. The purpose of the online access of the archival material was to allow the scientific research community easier access to many of the papers written during the infancy of the study of human genetics (Weiss and Lambert 2011). Some of these works have become classic papers in the subject (Fisher 1936), while others offer a glimpse into how racial prejudice can spill into scientific inquiry. One example of the latter is a series of articles written by the founder of the *Annals of Eugenics,* Karl Pearson, and his coauthor Margaret Moul entitled "The Problem of Alien Immigration into Great Britain, Illustrated by an Examination of Russian and Polish Jewish Children" (Pearson and Moul 1925a, b).

These series of articles are a fascinating read. While Pearson and Moul's purpose was to "discuss whether it is desirable in an already crowded country like Great Britain to permit indiscriminate immigration" (1925a, p. 2), the critical reader soon becomes suspicious that the authors' real aim was to prove the genetic inferiority of Jews, despite their claims to the contrary. Pearson was an unabashed supporter of eugenics (and makes this clear in the aforementioned articles) but was also an important mathematician and attempted to support all of his claims with scientifically rigorous data analysis.

The purpose of the current work is to investigate how Karl Pearson's desire to promote government mandated eugenics policies (Pearson and Moul 1925c) distorted his scientific practice and statistical inquiry specifically through the statistical design employed, analysis performed and inference made regarding these Jewish children that was subsequently published in the *Annals of Eugenics*. In this way the article is less of a historical inquiry into the nature of Pearson's biases (see Semmel 1958, MacKenzie 1979, and Schaffer 2008 for more detail regarding Pearson's philosophy) and more of a scientific critique through the eyes of ethical statistical practice. This article serves as both a tale of science gone wrong and a warning to scientists today that even the most talented among us can allow our personal opinions to skew our analyses.

We begin by giving some background on Karl Pearson's contributions as a mathematician, his stance on eugenics and British nationalism, and his role as the founder and editor of the *Annals of Eugenics*. We then give some necessary background on the study that is referred to in Pearson and Moul's work. Finally, we investigate the study design and two particular sections of the first article in the series (1925a) and make a few comments regarding the conclusion statements made in the second article (1925b). The validity of the analyses in these sections is questioned in light of current standards for ethical statistical practice.

Karl Pearson, Mathematician and National Eugenist

Karl Pearson is commonly considered to be one of the founders of modern statistical science. His formal education began at age 15 (in 1872) at University College

School. He later attended King's college at Cambridge where he won the Third Wrangler position of the Mathematical Tripos in 1878 (Semmel 1958). His initial work after his formal education was as an author, lecturer, and lawyer. However, in 1884 he was offered a position at University College, London in applied mathematics, where he taught mathematics to young engineering students (Porter 2004). Pearson would be associated with University College in various capacities throughout his career.

Pearson's mathematical contributions are vast. In 1892 he wrote *The Grammar of Science*, a famous work covering many scientific themes. He is credited with being one of the first mathematicians to truly consider data as essential in scientific inquiry. Instead of only developing new probability theory, Pearson used this theory as a tool with actual data (Walker 1958). He is well-known for his work in various measures of correlation; perhaps the most widely used is Pearson's product-moment correlation coefficient. Many of his statistical procedures are still in use today, such as Pearson's Chi squared goodness of fit test (Plackett 1983). In 1901 Pearson established the journal *Biometrika*, which remains in print today and is considered to be one of the most important statistical journals.

According to Porter, Pearson "wanted the scientific method to become part of shared wisdom and a shared ethic, so that science could help to rebuild that sense of cultural unity" (2004, p. 124). But Pearson had very developed ideas of what a shared wisdom and ethic should be. He was a British nationalist, politically concerned with advancing the British people without regard to other nationalities (MacKenzie 1979). He was a eugenist as well, and thus it was quite natural for Pearson to consider the *genetic* advancement of the British people. Pearson states, "The student of national eugenics desires in every way to improve and strengthen his own nation. He would do this by intra-national selection for parentage, and by the admission wherever and whenever possible of superior brains and muscles into his own country" (1925a, p. 6). One way to influence this advancement was through immigration policy. Pearson and Moul felt that the important and lasting characteristics of immigrants were inherently genetic and perhaps tied to race (Schaffer 2008). They state that, in regards to immigration policy, it "may be unwise to disregard race entirely when a standard of physique and intelligence has been established" (1925a, p. 6).

One of Pearson's goals was to apply numerical and statistical rigor to the sociological investigation of eugenics. Pearson did not feel there was a publication outlet for the results of investigations into the "racial problems in man" (Pearson and Moul 1925c, p. 1). The *Annals of Eugenics* was created specifically to address this problem and to devote a journal to the study of eugenics. Pearson felt strongly that eugenics would soon be shaping national social policy in such a way as to advance the genetic composition of future generations (Pearson and Moul 1925c). However, only these papers focused on immigration and written by Pearson himself would investigate specifically grounds for racial discrimination. The future papers published in the *Annals of Euguenics* focused on the study of inherited genetic traits. Weiss and Lambert claim this failure in the stated purpose of the journal was because of the difficulty of accurately quantifying racial characteristics (2011). However, this difficulty was not entirely recognized at the time. Two reviews of the

first issue of the journal, by S. J. Holmes and E. A. Hooton, were both favorable to the effort Pearson had put forth in adding statistical rigor to the study of eugenics. Hooton felt that racial characteristics were purely physical and could therefore be measured, and he applauded Pearson's work in developing statistical tools to study these traits (1926a). His review of the journal in particular holds up the *Annals of Eugenics* as "an indispensable source of information and inspiration for all workers who endeavor to apply exact mathematical methods of treatment to anthropological data" (1926b, p. 551). Holmes agreed with Pearson's concern that there were, until the *Annals*, no outlets for mathematical papers on the topic of eugenics. He defended Pearson's claimed lack of prejudice against the Jewish people and noted the authors' "reasonable position" that immigration policy should be set so that only immigrants who can better the nation's genetic makeup should be allowed into Britain (1926, p. 232). We refer the interested reader to the works of Pearson (1938), Eisenhart (1974), Norton (1978) and Porter (2004) for more detail on the contributions and political ideals of Karl Pearson.

Background of the Articles

Karl Pearson worked closely for many years with the renowned biologist Francis Galton. In 1906 Galton established the Galton Laboratory at University College, London with the intent of using the laboratory to further the study of human genetics (and in particular, eugenics) (Porter 2004, p. 278). In 1910, Pearson and Moul, as agents of the Galton Laboratory, carried out a large study involving hundreds of Polish and Russian Jewish schoolchildren attending the Jews' Free School in East London (Schaffer 2008). The Jews' Free School educated many of the immigrant Jewish children that had been migrating to London as a result of persecution in the late nineteenth and early twentieth centuries (Holmes 1995). This study included various examinations of the children's intelligence, health, and cleanliness, as well as the literacy of their parents.

The articles by Pearson and Moul published from 1925 to 1928 in the *Annals of Eugenics* outline the results of this study as well as statistical analyses and conclusions. The authors state that an immigration standard *could* be set that allows entry only to those individuals whose physique and intelligence are at an acceptable level. They then argue that there are two problems with this approach. One, an immigrant might not acclimate well to the British climate. Two, an immigrant might not assimilate into the British culture, which is desirable from Pearson's point of view. Therefore, the authors argue, race itself may be used as a criterion for immigration. The purpose of their study, as stated, was to investigate if differences in intelligence, health, and cleanliness could be attributed to race, and therefore to show that race would be a suitable choice for an immigration criterion.

Pearson and Moul felt they had accomplished this goal, as is evidenced in the following quote from the conclusion in Part II. "Taken on the average, and regarding both sexes, this alien Jewish population is somewhat inferior physically and mentally to the native population…But we have to face the facts; we know and admit that some of the children of these alien Jews from the academic standpoint

have done brilliantly, whether they have the staying powers of the native race is another question. No breeder of cattle, however, would purchase an entire herd because he anticipated finding one or two fine specimens included in it…" (1925b, pp. 126–127).

Experimental Purpose and Design

Before considering individual topics covered in the Pearson and Moul article, various aspects of the study itself are critiqued in light of moral and scientific ethical obligations. We point the interested reader to the American Statistical Association's (ASA) *Ethical Guidelines for Statistical Practice* (ASA 1999). The Preamble states the following: "statistical tools and methods, as with many other technologies, can be employed either for social good or evil…where the end purpose of a statistical application is itself morally reprehensible, statistical professionalism ceases to have ethical worth" (p. 3). We agree with this statement, and Pearson and Moul's work against the Jewish people (Levy 2004) in particular is certainly morally reprehensible. They were not alone, however, in considering a racial criterion for immigration. In the United States, for example, the purpose of the 1965 immigration and nationality act was to abolish an immigration criterion that was essentially race-based (Keely 1971).

In addition, various aspects of the data gathering process were ethically and scientifically questionable. It seems that the permission to test these school children was granted by the headmaster of the school, not the children and not their parents. There isn't any evidence in the paper that parents or children were given the option to refuse these tests. We recognize that guidelines for the treatment of human subjects were not in place in 1925; it wasn't until the Nuremburg code of 1947 that the rights of experimental subjects were given strong protection. However, we feel that Pearson and Moul's treatment of these children is not to be excused because of a lack of written, professional mandates. For example, it seems at least one eye examination was either painful or simply terrifying to some children. Pearson and Moul write that "examination of the accurate character we attempted took a long time, and the girls-even some of the elder ones-became nervous and there were even hysterical threatenings" (1925a, p. 9).

Pearson and Moul sampled children from only one school, and they do not indicate that they made any effort to sample Jewish children from any other school or part of Britain. They furthermore claim that the 600 children from this one school provided "a fair random sample of the population" (p. 9). This seems to be a very strong and cavalier statement, considering that the population they wished to make inference to was the entire Jewish population.

The medical examinations were performed by two people who were appointed by the Galton Laboratory and the ophthalmologist was also appointed by the Laboratory. Pearson and Moul claim that since these various "sympathetic" examiners were mostly Jews, "we can safely say that while it has probably been impossible to avoid errors, they are not biased errors." (1925a, p. 9). Yet there is no reason to believe an English born Jew could not be biased against these Russian and

Polish immigrant Jews. And there is certainly no mention of the training these examiners received with respect to minimizing bias in measurements.

Furthermore, they only observed Jewish children. Comparisons are made throughout the paper to Gentile children, but these data were gathered by other scientists at other periods of time, with up to 5–10 years difference in data sets. These comparisons are highly suspect, and it is unknown as to why Pearson and Moul did not choose to examine any non-Jewish children from other parts of London in tandem. In the multiple tables showing comparative data between Jewish and Gentile children, Pearson and Moul make a number of inferential claims about the population of alien Jewish children based on tables of comparative averages or percentiles. However, they rarely disclose the sample sizes and often give no test or correlation results, only observational inference. So there is no indication that these sample differences actually represent real differences not due simply to chance.

It would also seem that included in this sample are many children who are siblings. Pearson and Moul make no mention of this, and certainly do not control for the reduction in estimator variability that would potentially result from multiple siblings. Perhaps this mistake would have been caught by a reviewer, but as Pearson was the editor and founder of the *Annals of Eugenics*, we are not at all convinced a peer-review process was executed for these articles.

Finally, we note the offensive and inappropriate language often taken by the authors with respect to the Jewish immigrants. They recognized the oppression these Jews had been under in Russia and Poland and make the conclusion that "Such a treatment does not necessarily leave the best elements of a race surviving" (1925a, p. 8). In regards to the literacy of the parents of these children they write "…it follows [their results] that of the fathers 64.3 % and of the mother 82 % fail to have the knowledge of English requisite to make their generation effective citizens of this country" (1925a, p. 10). In regards to the history of the Jewish people in Poland they write, "The Jews as immigrants were never absorbed, and they did not become an integral portion of the people among whom they lived. That essential feature of Jewish immigration—a feature which differentiates the Jewish largely from other immigrant races, and widely from the other religious sects of this country—must always be kept in mind" (1925a, p. 18). It is hard to escape the suspicion that Pearson and Moul were *not* simply addressing a question of appropriate immigration criterion, but were in fact, aiming to make a case for the disallowance of the immigration of Jews into Great Britain.

Relationship of Parental Literacy to the Intelligence of the Children

One of the first investigations Pearson and Moul describe is the assessment of any relationship that might exist between the literacy of the children's parents and the intelligence of the children themselves. We consider first the analysis methods used. The children's intelligence was measured by their own teachers using a scale of Pearson's design. He states that "it has been shown that this scale adequately applied is…sensibly independent of age, i.e. it is a measure of innate mental power, not of acquired knowledge" (1925a, p. 10). Yet Pearson gives no references to the

reader as to how this scale has been shown to be a true measure of innate mental power. Pearson and Moul report the following contingency table (we give the boys only) (Table 1).

We consider a graphical representation of these data in Fig. 1. The categories for intelligence are determined by Pearson's intelligence scale. The barplot below does not show any *obvious* change in the distribution of intelligence scores based on the number of literate parents.

Pearson and Moul computed correlation ratios, but were not satisfied with the lack of significant results, so they collapsed the categories and reported the Tables 2 and 3.

We were able to reproduce the tetrachoric correlation reported by Pearson and Moul using R's (R Development Core Team 2011) polychor function for Table 3,

Table 1 Reproduction of Pearson and Moul's table III—Boys (1925a, p. 11)

Parents	Very able	Capable	Intelligent	Slow intelligent	Dull	Very dull	Mentally defective
Both literate	9	40	115	92	26	4	0
One literate	8	32	80	83	30	7	1
Both illiterate	1	8	26	33	11	1	0

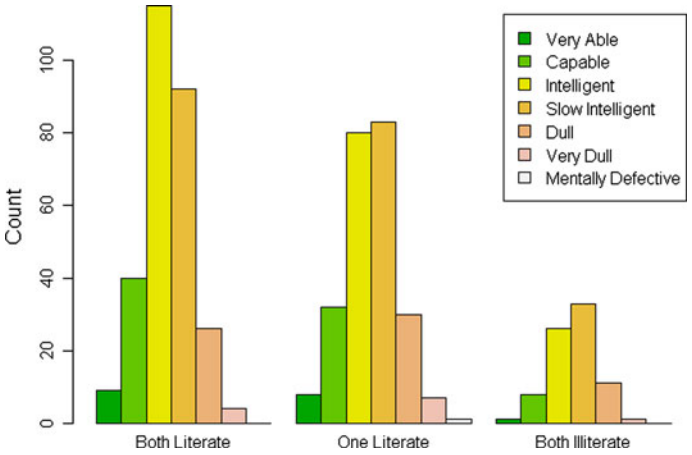

Fig. 1 Graphical representation of the data in Table 1

Table 2 Reproduction of Pearson and Moul's table "Daughter's intelligence" (1925a, p. 12)

Parents' literacy	Girls	
	Intelligent and over	Slow and under
Both literate	94	169
One or both illiterate	90	206

Table 3 Reproduction of Pearson and Moul's table "Son's Intelligence" (1925a, p. 12)

Parents' literacy	Boys	
	Intelligent and over	Slow and under
Both literate	164	122
One or both illiterate	155	166

but not for Table 2, and question whether the reported correlation coefficient reported is an error. We examined these data using another method development by Pearson to examine dependent relationships, the Pearson Chi squared test for independence. For the data in Table 2, the obtained p value is 0.21 and for Table 3, it is 0.03. The conclusions from these two tests would be that there is no evidence for a relationship between parental literacy and intelligence for the girls, and some evidence of a relationship at the $\alpha = 0.05$ level for the boys. Pearson and Moul however, state that, for the girls, the tetrachoric correlation "cannot be asserted to be definitely significant" (our analysis shows no evidence at all for a relationship, and we suspect the correlation reported is an error). For the boys, they state their tetrachoric correlation "is most probably significant and therefore confirms the not definitely significant results from the girls" (1925a, p. 12). Not only do we question their determination of significance, but it is certainly not the case that a significant result from one test and set of data can confirm the significance of a *non-significant* result from another test and a different set of data!

Pearson and Moul conclude that they have found a relationship between the literacy of the parents and the intelligence of their children, but it is slight. They state, "We must accordingly conclude that to admit only the literate immigrants would not raise appreciably the offspring population, and therefore since intelligence is strongly inherited the illiteracy of the parents is rather a measure of their want of education than their want of intelligence" (1925a, p. 13). However, as is evidenced in much of their work, Pearson and Moul conclude the reporting of non-compelling results with statements that give additional reasons why, despite their non-significant results, these immigrants should *still* not be admitted. The previous comment by Pearson and Moul was immediately followed by this statement: "The refusal to admit illiterates must be regarded from the standpoint of the parental rather than the offspring generation. Can an illiterate man make a good citizen?" (p. 13).

Resistance to Tuberculosis

In addition to investigations of intelligence, Pearson and Moul also considered various measures of physical health of the Jewish children and compared these measures to those taken previously (not by Pearson and Moul) on British Gentile children. One of these investigations focused on the health of the cervical (neck) glands, supposedly for the purpose of considering a resistance to tuberculosis. They observed that Jewish children have healthier glands than Gentile children, and this might suggest that Jews are more resistant to tuberculosis. We note that Pearson and

Moul give no indication as to where the "Average Gentiles" data was obtained that was used in comparison to the Jewish children (p. 42).

Pearson and Moul's acknowledgement of the relatively healthier glands of the Jewish children is immediately followed by a lengthy argument for the higher incidence of tuberculosis disease among the alien Jews than among the Gentile children. It is this argument we consider.

Pearson and Moul begin by citing a study conducted by William Moses Feldman which examined tuberculosis cases from 1901 to 1906 occurring in various European towns. Pearson and Moul cite the London data given by Feldman for the mortality rate from tuberculosis per 10,000. For Jews the rate was 13.3 and for Gentiles, 17.9. Pearson and Moul immediately question the use of these data as evidence for a reduced tuberculosis mortality rate for Jews. As Pearson and Moul do not seem to question any of their own data or inference that supports, they claim, the inferiority of Jewish peoples, we view this argument as motivated by the bias Pearson and Moul have towards Jewish immigrants.

Their argument is constructed as follows. First, they question if the death-rates had been standardized to account for distributional age differences among the populations. They state that "true standardization would probably cause considerable changes" (p. 42). Yet they give no evidence as to whether or not the data have been standardized, much less any idea as to the size or direction of the effect of standardization on any inference. Second, they challenge the appropriateness of the use of Feldman's sample (the data was only from Whitechapel, a borough of London on the east end) to make inference to all Jews in London. We note, of course, that Pearson and Moul use a sample of Jewish children from one school to make inference to all Jewish peoples.

Pearson and Moul then choose to consider death-rates from tuberculosis from two Polish towns during the years of 1896–1902. These data provided a tuberculosis mortality rate for Jews of 20.5 per 10,000 from one town and 30.6 per 10,000 from another town. These rates are much higher than Feldman's reported rates. However, mortality rates from any communicable disease are related to exposure, which is relative to a particular time and place. Pearson and Moul chose data much older than Feldman's and from a different part of the world, yet they do not compare these data with that of non-Jewish people in the same time and geographic location.

This challenge did not go unnoticed by Feldman. His response was published in the British Medical Journal in 1926. Feldman was, in part, responding to the following statement:

> It is quite possible that the Jews are more anxious and visit the dispensary with slighter symptoms… The lower mortality may not be due to a lower incidence rate, but to what has been stated as a fact, that the disease runs a more chronic and less fatal course in Jews than Gentiles. A chronically affected population may be *less efficient* [emphasis added] than one not so affected but having a higher death-rate. (Pearson and Moul 1925a, pp. 42–43)

Feldman indicates that this is not the correct conclusion. He claims that the mortality rate from tuberculosis is lower among all Jews. He further states that "the disease tends to cure itself" among Jews, as evidenced by data from a study at a

sanatorium comparing the percentage of cures from tuberculosis between Jews and Gentiles, which showed that "the percentage of cures is much higher amongst the Jews than amongst the Gentiles" (Feldman 1926). Feldman notes that Pearson and Moul did not choose to reference any of the evidence that was in favor of a stronger Jewish resistance to tuberculosis.

Finally, Pearson and Moul introduced aggregated data from an unknown number of School Medical Officers' Reports. The Table 4 is a reproduction of the data.

From these data Pearson and Moul (amazingly) state that "these statistics seem to indicate a greater prevalence of tuberculosis disease among the alien Jew than among the average children of London elementary schools" (1925a, p. 43). However, these data should not even be compared. There is no indication as to the particular tuberculosis diagnosis for the Gentile children (active or quiescent/doubtful). If the Gentile data only contained "active" tuberculosis diagnoses, then the alien Jewish children would have a *smaller* percentage suffering from active tuberculosis for both boys and girls. The conclusions made by Pearson and Moul are wholly inappropriate given the data and they violate many current ethical guidelines for statistical inference (ASA 1999, pp. 3, 4, 6).

Historical Responses to the Work of Pearson and Moul

Pearson and Moul's work did not go unnoticed amongst the Jews of Great Britain. *The Jewish Chronicle* was (and still is) in print at the time of the publication of Pearson and Moul's work and there was much published regarding their conclusions. Prior to the publication in the *Annals of Eugenics*, Pearson had presented his work as a talk entitled "The Alien Jewish Child" at the annual conference of the Union of Jewish Literary Societies, held at Jews' College ("The Alien," 1922). A Mr. E. I. Spiers subsequently wrote a letter to the editor of *The Jewish Chronicle* questioning Pearson's "far-fetched inference" regarding the potential racial intermixture of Polish and Russian Jews using the cephalic index. He was also concerned about Pearson and Moul's experimental design, writing that "Professor Pearson's investigations appeared to be dotted with flaws, partly due to the fact that they suffered from over-localization, or restriction to a fragmentary area of Jewish children, and were not spread over a larger section of alien Jewish children" (Spiers 1922).

Table 4 Reproduction of Pearson and Moul's table XLIII (1925a, p. 43)

	Boys			Girls		
	Entrants (%)	8–9 years (%)	12–13 years (%)	Entrants (%)	8–9 years (%)	12–13 years (%)
Gentiles, average, all cases	0.60	0.62	0.74	0.52	0.62	0.68
Alien Jews, active		1.62			1.37	
Alien Jews, quiescent and doubtful		1.13			1.37	

Another letter to the editor was written by a Mr. Samuel H. Dainow, in which he called into question (quoting multiple scholarly sources) the validity of the use of intelligence scales with children who did not speak English as their primary language in the home (Dainow 1922). The Headmaster of the Jews' Free School, Mr. L. G. Bowman, who had given permission to Pearson and Moul to conduct the investigation using the children of the school, read a paper at the War Memorial Conference in 1922 that refuted Pearson and Moul's claims. He stated in his reading that he had no idea that the study was in regards to questions of immigration policy ("Problems of Jewish," 1922).

An article written by a well-known rabbi not only expressed great concern regarding the validity of the data gathered by Pearson and Moul, but also questioned, albeit very carefully, their motives:

> [Pearson and Moul] say that they are not against *all* immigration into this country, and they deny that they are criticizing the alien Jewish immigration simply because it is *Jewish.* There is, therefore, no need to impute to them the slightest anti-Jewish prejudice. But, nevertheless, they are not altogether free from the human liability to pre-judgment, or, what psychologists call "apperception," that selective activity on the part of the mind which tends eagerly to find what it seeks (Levy 1923).

In 1927 a paper by Mary Davies and A. G. Hughes challenged the findings of Pearson and Moul (in regards to intelligence, specifically) by analyzing data gathered from multiple schools that contained both Jewish and non-Jewish children. The authors themselves were not Jewish, but their results were a refutation of the conclusions made by Pearson and Moul (Davies and Hughes 1927). A Dr. E. Bernstein hailed Davies' and Hughes' work as "immensely superior to that employed in Professor Pearson's work" (Bernstein 1927). Bernstein noted the improved sampling methodology and, in particular, the more "refined" intelligence tests.

This subsequent study by Davis and Hughes attempted to control for many of the sampling biases seemingly ignored by Pearson and Moul. They sampled students from multiple schools in order to control for variation due to different education styles. They sampled from schools containing both Jews and non-Jews and chose these schools so that the students were of a similar socio-economic class. In fact, Davies and Hughes noted that the majority of children examined by Pearson and Moul were recent immigrants, with many actually born abroad. In Pearson's work these children were compared to non-immigrant Gentile children, often without comment as to differing social and economic class status. Furthermore, Davies and Hughes rated the children's intelligence based on the Northumberland standardized tests, which are not reliant on teacher ratings. They did, however, compare their intelligence results with those obtained using Pearson's Biometric Laboratory scale and found that their two sets of intelligence results confirmed each other in rating Jewish children as more intelligent (in many categories) than non-Jewish children. Their conclusion was that teacher's estimates of intelligence are not transferable across schools. In other words, Pearson should not have compared the intelligence scores of the Jewish children in the Jews' free school with the scores obtained for

non-Jews in other schools. This critique of the Biometric scale was echoed approximately 10 years later when a review was performed by Moshe Brill of Ohio State University of the comparative studies of the intelligence of Jews and non-Jews (1936). This review contained both Pearson and Moul's paper as well as that of Davies and Hughes. Below is Brill's critique of Pearson and Moul's study.

> The main criticisms of this study are: (1) The ratings were made by different teachers. (2) Teachers' ratings of the intelligence of foreign children are notoriously unreliable. (3) The Jewish and non-Jewish children were not of the same socio-economic status. Only 6 per cent of the fathers of the Jewish children, for instance, were native-born or naturalized, and 71.9 per cent of them were reported to be engaged in occupations of low-grade labor. (4) Pearson's contention that the father's occupation is no criterion of the innate intelligence of the child is unfounded. (5) The reliability and validity of the Biometric Scale for measurement of intelligence has not been established (p. 332).

His critiques of Davies and Hughes' study are only that "the samples in most age groups were too small to permit generalizations" (which Davies and Hughes consider in their paper) and "the factor of language handicap was not ruled out" (p. 326). It would seem that even given the standards of the day, Pearson and Moul's study lacked the rigor necessary for the strong conclusions of Jewish inferiority they make throughout their work.

Conclusions

Throughout these series of articles Pearson and Moul adopt what we consider to be a tone of scientific superiority. They display significant amounts of data, perform a large number of statistical tests and methods, and make numerous inferences. Many of these inferences are simply not warranted given the data that was reported. We find that Pearson and Moul's statistical analysis and scientific method leaves much to be desired. They used suspicious sampling methods and data acquisition methods that could have introduced bias and their comparison datasets were often not adequately explained (sources, sample sizes, etc.). They even seemed to ignore a seemingly obvious problem in examining siblings.

We recognize that not all current methods and standards of practice were in use in the early twentieth century. Yet the most egregious offenses and ethical violations in Pearson and Moul's work cannot be excused because the science of statistics was still in its infancy. Pearson and Moul used derogatory and insensitive language towards the Jewish population and only gave discussion to reasons why Jewish immigration should be halted, despite the data, much of which they did not gather nor provide sufficient documentation for. While Pearson may not have been a firm anti-Semite, he was an elitist. He viewed society as divided into classes, and himself as in the class of the educated and learned person. He viewed this class as more important than the working class and in fact, viewed the "Proletariat pure and simple" as a threat to social stability (MacKenzie 1979, p. 130). Pearson states that

"No more than there is equality between man and man of the same nation is there equality between race and race" (Pearson and Moul 1925c, p. 3).

Because of this elitism, it seems that the Jews' free school might have been chosen so that it would be easier to establish racial inferiority. Regardless of whether or not this was the intent, Pearson seemed to ignore the obvious comparability problems regarding social-economic class, etc., in ways he did not in other studies. For example, Stigler discusses at length a study Pearson performed regarding the potential effects of alcoholism on children in an effort to scientifically address claims made by the temperance movement. In this study it seemed that Pearson took pains to ensure the data was representative and were not gathered with the goal of addressing questions of alcoholism (1999, Ch. 1).

Yet when studying Jewish immigrants, Pearson and Moul exhibited in their work an amazing lack of caution and humility in their inference, especially considering the ramifications of their recommendations. We hope this critique will shed light on how science can be performed poorly, especially when a particular end is sought. This example should serve as another reminder that as scientists, we are not without personal opinions. We must be diligent both ourselves and within our peer-review community to report our findings respectfully, transparently and with due attention to potential errors. We end with a quote from Pearson and Moul's conclusion in part II:

> Nor do we criticize the alien Jewish immigration simply because it is Jewish; we took the alien Jews to study, because they were the chief immigrants of that day and material was readily available. We know what the world owes to Jewish philosophers, to Jewish musicians and to Jewish writers. We should have stood by Reuchlin when he cried: "Print rather than burn the Jewish books." Our chief fear in checking indiscriminate Jewish immigration is not that Britain may lose a supply of cheap labour, but that we may exclude a future Spinoza, a Mendelssohn, a Heine or an Einstein. Yet approaching the problem sympathetically and as we hope without bias, we cannot see that unrestricted immigration has been an advantage to this country. At first sight it seems a hard judgment. If we accept some of these Polish and Russian Jews as reaching a proper immigration standard, are we to leave the majority of them to suffer the hardships, possibly the tortures they have met with in Eastern Europe? Assuredly this is not the only alternative. There are lands less crowded than our own; there are populations physically and mentally below the level of these alien Jews. If Palestine be such a land, then the associations of the past are an invaluable aid. But if it be such a land the Jewish immigrant must go as the Danes went to Yorkshire, with a spade in one hand and a weapon in the other. For it is not town-workers, but fruit and corn-growers that are needed, and the Arab race will not indefinitely allow soil to the man who cannot defend it and himself. If Palestine fulfill not our immigration conditions, then surely the wisdom of our statesmen might find another home for the bulk of these persecuted Jews than the crowded cities of Western Europe? (1925b, p. 127)

References

ASA [American Statistical Association]. (1999). Ethical guidelines for statistical practice. http://www.amstat.org/about/ethicalguidelines.cfm. Accessed July 1 2012.

Bernstein, E. (1927). *The mentality of the Jewish child* (p. 14). UK: The Jewish Chronicle.

Brill, M. (1936). Studies of Jewish and non-Jewish intelligence. *Journal of Educational Psychology,27*(5), 331–352.

Dainow, S. H. (1922). *Handicap to yiddish-speaking children* (p. 17). UK: The Jewish Chronicle.

Davies, M., & Hughes, A. G. (1927). An investigation into the comparative intelligence and attainments of Jewish and non-Jewish school children. *British Journal of Psychology,18*(2), 134–146.

Eisenhart, C. (1974). Karl Pearson. In C. C. Gillispie (Ed.), *Dictionary of scientific biography* (pp. 447–473). NewYork: Scribner.

Feldman, W. (1926). Alien immigration into great Britain. *British Medical Journal,1*(3395), 166–167.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics,7*(2), 179–188.

Holmes, S. J. (1926). Annals of eugenics: A journal for the scientific study of racial problems. *Science,63*(1626), 232–233.

Holmes, C. (1995). Jewish economic and refugee migrations, 1880–1950. In R. Cohen (Ed.), *The Cambridge survey of world migration* (pp. 148–153). Cambridge: University Press.

Hooton, E. A. (1926a). Methods of racial analysis. *Science,63*(1621), 75–81.

Hooton, E. A. (1926b). Review: Annals of eugenics. A journal for the scientific study of racial problems by Karl Pearson. *American Anthropologist,28*, 551–553.

Keely, C. B. (1971). Effects of the immigration act of 1965 on selected population characteristics of immigrants to the United States. *Demography,8*(2), 157–169.

Levy, S. (1923). *Children of Jewish immigrants* (p. viii). UK: The Jewish Chronicle.

Levy, D. M. (2004). Statistical prejudice: From eugenics to immigrants. *European Journal of Political Economy,20*, 5–22.

MacKenzie, D. (1979). Karl Pearson and the professional middle class. *Annals of Science,36*(2), 125–143.

Norton, B. J. (1978). Karl Pearson and statistics: The social origins of scientific innovation. *Social Studies of Science,8*(1), 3–34.

Pearson, E. S. (1938). Karl Pearson: An appreciation of some aspects of his life and work. *Biometrika,28*(3), 193–257.

Pearson, K., & Moul, M. (1925a). The problem of alien immigration into Great Britain, illustrated by an examination of Russian and Polish Jewish children: Part I. *Annals of Eugenics,1*(1), 5–54.

Pearson, K., & Moul, M. (1925b). The problem of alien immigration into Great Britain, illustrated by an examination of Russian and Polish Jewish children: Part II. *Annals of Eugenics,1*(2), 56–127.

Pearson, K., & Moul, M. (1925c). Foreward. *Annals of Eugenics,1*(1), 1–4.

Plackett, R. L. (1983). Karl Pearson and the chi squared test. *International Statistical Review,51*(1), 59–72.

Porter, T. M. (2004). *Karl Pearson: The scientific life in a statistical age*. Princeton, NJ: Princeton University Press.

Problems of Jewish education (1922). *The Jewish Chronicle*, p. 18.

R Development Core Team (2011), *R: A Language and environment for statistical computing*. Vienna, Austria. ISBN 3-900051-07-0; URL http://www.R-project.org/.

Schaffer, G. (2008). Assets or 'aliens'? race science and the analysis of Jewish intelligence in inter-war Britain. *Patterns of Prejudice,42*(2), 191–207.

Semmel, B. (1958). Karl Pearson: Socialist and darwinist. *The British Journal of Sociology,9*(2), 111–125.

Spiers, E. I. (1922). *Dotted with flaws* (p. 30). UK: The Jewish Chronicle.

Stigler, S. M. (1999). *Statistics on the table*. Cambridge, MA: Harvard University Press.

The Alien Jewish Child (1922, May 12). *The Jewish Chronicle*, p 15.

Walker, H. M. (1958). The contributions of Karl Pearson. *Journal of the American Statistical Association,53*(281), 11–22.

Weiss, K. M., & Lambert, B. W. (2011). When the time seems ripe: Eugenics, the *annals*, and the subtle persistence of typological thinking. *Annals of Human Genetics,75*, 334–343.