13. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods* 2010;15:309–34.
14. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol* 2010;172:1339–48.
15. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* 2010;21:540–51.
16. VanderWeele TJ. Mediation analysis with multiple versions of the mediator. *Epidemiology* 2012;23:454–63.

Commentary: Smoking, birthweight and mortality: Jacob Yerushalmy on self-selection and the pitfalls of causal inference

Mark Parascandola

Tobacco Control Research Branch, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892-9761, USA. E-mail: paramark@mail.nih.gov

International Journal of Epidemiology, 2014, 1373–1377
doi: 10.1093/ije/dyu163

In the television show *Mad Men*, which goes to great lengths to maintain historical accuracy as to its 1960s setting, a visibly pregnant Betty Draper is shown smoking in the maternity ward. Little research had been conducted at the time on the potential effects of smoking during pregnancy.¹ Jacob Yerushalmy's 1971 paper appeared as a wave of new findings pointed towards the detrimental impact of maternal smoking on the developing fetus. But the paper goes beyond merely reporting study results. Yerushalmy methodically sifts through the data from multiple angles, testing a range of potential hypotheses, while constructing an argument about the perils of causal inference from observational studies. The paper was one of a series, part of a discussion that persisted from the 1950s to the 1970s and possibly beyond. Understanding this historical context is essential to appreciating Yerushalmy's contributions.

On epidemiological methods and causal inference

Yerushalmy (1904–73) was born and raised in what is today known as Israel, until he came to the USA in 1924 to attend university, studying mathematics at Johns Hopkins University. Later, as an instructor at Hopkins, home to Raymond Pearl and Wade Hampton Frost, he was exposed to the growing field of biostatistics. Yerushalmy held a series of posts as a statistician in the New York State Department of Health, the Public Health Service and the Children's Bureau of the Department of Labor, where he began studying child development.² In 1947 Yerushalmy

moved to the University of California at Berkeley to found a new department of biostatistics. Beginning in 1959, he also led the Child Health and Development Studies (CHDS), a cooperative project with Kaiser Permanente Hospital. The CHDS was innovative in conducting long-term follow-up of participants in one of the first health maintenance organizations.³ This ongoing prospective study became the locus of Yerushalmy's work throughout the following decade.

During the 1950s, Yerushalmy became an active commentator on epidemiological methods and causal inference, bringing attention to sources of bias and flawed analyses. For example, along with Herman E Hilleboe of the New York State Department of Health, he criticized ecological studies comparing dietary fat and heart disease mortality across a selection of countries. They pointed out that the countries differed in many respects beyond the two variables of interest, and also that the seemingly arbitrary selection of countries for analysis tended to favour the result. Investigators should, they recommended, assess whether 'the association between two variables is in fact between the variables investigated and does not merely reflect relationships with a broader group, of which one or the other of the variables forms a part'.⁴

But most notable was the influential 1959 paper by Yerushalmy and Palmer, which urged that epidemiologists should seek to imitate 'the more rigorous methods long in use by bacteriologists'. Bacteriologists had a set of rules—Koch's postulates—for drawing aetiological conclusions about infectious agents. But these methods relied on identification of specific disease agents and laboratory

demonstrations. In studying environmental causes of cancer, epidemiologists were forced to rely on observational data on risk factors, such as cigarette smoke, in place of specific disease agents. Moreover, to mimic the experimental condition it was essential to make sure the two groups were identical in all possible respects except the variable under study. But this condition could not be assumed in non-randomized observational studies. The culprit here was self-selection, as observational study participants are not assigned to conditions of smoking or not smoking, but make this choice for themselves. '[S]moking, like volunteering, may represent an index which differentiates the two groups in many aspects of mode of life and perhaps also on constitutional grounds', they wrote.⁵

To help rule out non-causal spurious associations, Yerushalmy urged that investigators should test for the 'specificity' of an association. That is, our confidence that a causal relationship is not spurious should be greater when a cause is associated with fewer effects (and vice versa). The rationale here is that a factor which appears to be associated with many dissimilar outcomes may reflect some form of study bias. For example, statistician Joseph Berkson argued that the fact that most (88.5%) of the excess deaths among smokers in the Hammond and Horn study were not from lung cancer, but from various other causes, indicated the presence of some sort of selection bias.⁶

Those epidemiologists who argued that the evidence was sufficient at the time to call cigarette smoking a cause of lung cancer, appealed to causal criteria also, but a different set of criteria. For example, the strength of an association was held to be important because a strong non-causal association required a strong confounder to explain it, and weak associations were more likely to be artefacts of selection bias. Such an obvious fact could hardly escape the attention of any conscientious investigator, the reasoning went.⁷ In contrast, Yerushalmy had rejected strength of association as a criterion for inference because its evaluation was necessarily subjective: 'There is no rational way to decide how large a difference there must be before we accept it as indicating a cause-effect relationship'.⁸ At the same time, the lack of specificity of smoking as a cause of lung cancer was not a source of worry, according to Abraham Lilienfeld, because the association between smoking and lung cancer in particular was so dramatic relative to other adverse effects.⁹

Smoking, low birthweight and mortality

During the 1950s, as concerns grew about the health effects of cigarette smoking, several studies appeared suggesting that smoking mothers were more likely to have

babies of lower birthweight compared with nonsmoking mothers. For example, a 1957 study by Simpson found that mothers who smoke during pregnancy have a much larger proportion of their births in the 'low birthweight' group (2500 g or less) than mothers who do not smoke. These results raised much concern because low birthweight infants were known to experience very high neonatal mortality.¹⁰ The 1964 report of the Surgeon General on Smoking and Health commented on these studies but acknowledged that the underlying mechanism remained unclear and, more importantly, that 'It is not known whether this decrease in birthweight has any influence on the biological fitness of the newborn'.¹¹

Yerushalmy was sceptical of these findings. He reasoned that if cigarette smoking was in fact deleterious to the developing fetus, through some as yet unknown mechanism, then one would expect the babies of smoking mothers to also experience higher mortality. But so far, the studies had only shown a difference in weight, not in mortality. In order to make sense of these findings, Yerushalmy argued for the importance of following the smoking and nonsmoking mothers prospectively and tracking infant survival. At the outset, he noted that such a study would also have important implications for 'the more general question of the validity of inferring causal relationships from observed associations'.¹² The questions of causal inference and smoking during pregnancy were closely intertwined from the start.

Yerushalmy turned to the CHDS to study the issue, using participants in the Kaiser Foundation Health Plan from the San Francisco—East Bay area. Because women in the plan were entitled to prenatal care, they tended to start medical care early in pregnancy. Upon entry into the study, a lengthy interview was conducted, including a detailed smoking history of both parents. Participation in the study was almost 100% and by 1 April 1963, after 3 years running, the study included over 10 000 pregnancies. Because most participants stayed in the plan, it was possible to follow them up and assess whether smoking habits changed during or after pregnancy and to track the health of the offspring over time.

The first report from the study, published in 1964, found unsurprisingly a greater proportion of low birthweight infants among smokers than nonsmokers. Additionally, smoking a greater number of cigarettes per day was associated with a greater probability of having a low birthweight baby, suggesting a dose-response relationship. But no difference in mortality was seen between infants of smoking vs nonsmoking mothers. Yerushalmy and colleagues delved further into the data to try to explain this conundrum. When looking only at lowbirth weight infants, they found surprisingly that the low birthweight

infants of smoking mothers survived considerably better than those of nonsmoking mothers. This appeared highly implausible. 'It is difficult to visualize a biological mechanism whereby mother's smoking is the cause for these phenomena—that it exerts a beneficial effect on the infant of 'low birthweight' which reduces markedly his risk of early death,' wrote Yerushalmy. Instead, Yerushalmy suggested that 'It is not the smoking but the smoker which may offer an explanation for the observed differences'. In other words, the smokers in the study may represent a different group of people whose reproductive experience would have been similar independently of whether or not they smoked. In this case, smoking is not a cause of the difference in birthweight, but is simply correlated with some other factors that are influential.

But the prevailing opinion was heading in the other direction. On 9 November 1968, the *British Medical Journal* published an editorial describing the substantial body of evidence linking smoking to low birthweight babies. Whereas they acknowledged that the mechanisms at work were unknown, they highlighted the rise in women smoking and in mortality from lung cancer among women: 'It seems that the time has come when women should be told frankly that if they smoke they not only put their own lives in jeopardy but, if they continue to do so during pregnancy, also expose their unborn infants to an unnecessary risk'.¹³

A prospective study of 2000 women, conducted by Scott Russell and colleagues at the University of Sheffield, had found that in addition to the expected differences in birthweight by smoking status, 'the percentage of unsuccessful pregnancies (abortion, stillbirth, neonatal death) was higher for smokers', even when controlling for education, social class, and other factors. This was the first large prospective study to have reported a difference in neonatal mortality. Extrapolating from the data, Russell went on to estimate that one out of every five unsuccessful pregnancies in women who smoked regularly would have been successful if the mother had not smoked during the pregnancy.¹⁴

On 11 January 1971, Surgeon General Jesse Steinfeld gave a preview of the latest smoking and health report to the National Interagency Council on Smoking and Health. He emphasized the growing impact of cigarette smoking on women and called for 'aggressive campaigns to reach every girl in the country with the health message on cigarette smoking'.¹⁵ Steinfeld described strong tobacco industry advertising campaigns targeted at women, noting that his office had counted 36 advertisements running in eight leading women's magazines at the time. And the new report, he promised, would contain data further strengthening the harmful effects of smoking during pregnancy.

The 1971 report released the following week did acknowledge that results regarding increased risk of spon-

taneous abortion, stillbirth and neonatal death were inconsistent. However, they cautioned that most of the results were based on retrospective studies among women who delivered in hospitals and infants whose names appeared on hospital newborn lists, which would tend to underrepresent women who had aborted. Thus, they placed greater emphasis on the prospective results. In addition to the Russell study, new data from the Second Report of the 1958 British Perinatal Mortality Survey published in 1969 found that 'the mortality in babies of smokers was significantly higher than in those of nonsmokers'. The report also noted Yerushalmy's unusual 1964 finding that low birthweight babies of smoking mothers seemed to fare better. But the authors countered that whereas the neonatal death rate for the small infants of smoking mothers appeared less than that for small infants of nonsmoking mothers, neither group could be considered 'healthy', having sharply elevated death rates compared with normal weight babies.¹⁶

Yerushalmy's 1971 paper reports on additional data from the ongoing CHDS, now including 13 000 pregnancies. Again, women who smoked had approximately twice as many low birthweight infants per 1000 single live births as women who did not smoke. And, again, the low birthweight infants of smoking mothers survived considerably better than low birthweight infants of nonsmoking mothers. Adding to the confusion, when considering the smoking behaviour of both parents, Yerushalmy found that 'The most vulnerable low-birth-weight infants are produced by couples in which the wife does not smoke and the husband smokes'.

The results continue to serve as a case study in causal inference. Observational studies 'remain the first line of attack in probing for causes of chronic diseases and conditions', so what is needed are 'auxiliary and complementary methods' to help overcome the built-in limitations of observational epidemiology. In this case, that means that testing alternative hypotheses, such as the biological or 'mode of life' differences between smokers and nonsmokers, may explain the results. Indeed, compared with nonsmokers, smokers in the CHDS study were less likely to use contraceptive methods, less likely to plan the pregnancy, more likely to drink hard liquor, beer and coffee and more likely to indulge in these behaviours to a greater extreme. Additionally, the age of menarche was significantly lower for women who subsequently became smokers than for nonsmokers. Although far from definitive, these findings seemed to suggest caution in implicating cigarette smoking.

The ongoing controversy

Following the release of the 1971 Surgeon General's report, Daniel Horn, head of the Clearinghouse on Smoking and

Health, was travelling the country delivering talks about the dangers of cigarette smoking during pregnancy. He told media outlets that the federal government was beginning a national crusade to 'give babies a fair chance'.¹⁷ In response to the Public Health Service campaign, the Tobacco Institute, representing the major cigarette manufacturers, developed a 16-page 'backgrounder' document for editors and journalists entitled *Smoking and Pregnancy: Is the Question Answered?*¹⁸

Yerushalmy's work gained media attention as well with headlines like 'Mothers needn't worry, smoking little risk to baby'.¹⁹ The following year, Yerushalmy conducted a further analysis, focusing on women who started smoking or quit smoking between multiple pregnancies. The results showed that the reproductive outcomes among those who changed their smoking status (from non-smoking to smoking or vice versa) were similar across time. In other words, the 'future smokers' produced low birthweight babies even before they started to smoke and the former smokers produced normal birthweight babies even after they quit. 'The evidence appears, therefore, to support the proposition that the incidence of low birth weight infants is due to the smoker and not the smoking.'²⁰ The paper received substantial news coverage, including a column in *Family Health Magazine* titled 'In defense of smoking moms'.²¹

The 1973 Surgeon General's report cited a 'strong, probably causal association' between mothers' smoking and infant mortality, estimating that 4600 stillbirths each year in USA were attributable to mothers' smoking in pregnancy and could be prevented.²² Shortly after the report's release, Senator Marlow W Cook of Kentucky gave a lengthy speech on the Senate floor, denouncing the report as 'dangerous misinformation'. In support, Cook read into the record a letter from Yerushalmy who wrote, at Cook's invitation, responding to critiques of his work in the reports of the Surgeon General. Yerushalmy reiterated points he had been making from the start, clarifying that his purpose was not to fight for a definitive conclusion but to maintain that the case should remain open.²³ Yerushalmy died later that year, aged 69.

It wasn't until 1980, with the report on *The Health Consequences of Smoking for Women*, that a definitive statement appeared in the reports of the Surgeon General regarding the influence of maternal smoking on neonatal mortality.²⁴ By this time, a substantial body of cohort data tracking birth outcomes was available. In response, the Tobacco Institute produced a 'backgrounder' for journalists on women and smoking. Discussion of maternal smoking, pregnancy and potential effects on children appeared up front, before cardiovascular disease or chronic obstructive pulmonary disease. The text continued to rely heavily on Yerushalmy's arguments.²⁵

Several broader changes also increased attention to environmental influences on pregnancy outcomes during this period. The thalidomide episode in the early 1960s raised alarm about how a substance ingested by a pregnant mother could seriously harm the developing fetus. The 1973, University of Washington studies of fetal alcohol syndrome further added to these concerns.²⁶ And the 1970s also brought a dramatic shift in the debate over smoking and health, acknowledging potential effects on non-smokers, including children and the developing fetus.

Yerushalmy's work with the CHDS was primarily funded by the National Institutes of Health, but there is limited evidence that he received some consulting fees from the tobacco industry, including support for a pilot survey in collaboration with the government of Israel.²⁷ He also sought funding from the Council for Tobacco Research for a large prospective study in 1972, but it appears this proposal was declined.²⁸ However, after Yerushalmy's death, statistician Richard J Hickey and colleagues at the Wharton School continued work along similar lines, with support from the Council, into the 1980s. When Congressional hearings were held in 1982 to modify the standard cigarette warning labels, including the addition of a statement about smoking during pregnancy, Hickey testified against the bill, citing Yerushalmy on the self-selection problem in studies of smoking and pregnancy.²⁹ But the Comprehensive Smoking Education Act of 1984 (Public Law 98-474) eventually mandated four new, more specific health warnings on all cigarette packages and advertisements, including the statement that 'Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight'.

Why did Yerushalmy argue so vehemently and persistently on the issue? Yerushalmy was not so much arguing for a particular hypothesis as he was against a rush to judgment. Coming from an earlier generation of biostatisticians, he likely feared a backlash against the discipline's hard-won authority. In this sense, his motivation was likely similar to that of Ronald A Fisher and other biostatisticians who ended up on the 'wrong' side of the debate over smoking and health.³⁰ As outmoded as the constitutional hypothesis may seem today with regard to smoking and lung cancer, substantial efforts were devoted at the time to characterizing smoking and nonsmoking phenotypes.^{31,32} These exchanges did help advance the field of epidemiology by forcing those on both sides of the issue to further develop their arguments and identify methods to address potential weaknesses.³³

Possibly the most underappreciated lesson of Yerushalmy's work relates to the use of causal criteria. Yerushalmy was not opposed to such criteria; in fact, he argued they were a necessary adjunct to observational epidemiology. But, at the same time, he recognized that the

criteria were subject to judgment and interpretation when applied in practice. Yerushalmy gave priority to specificity and coherence while others emphasized strength and consistency. And how one interprets 'strength' or 'specificity' may also be subjective. Epidemiological review papers today often include a passing nod to causal criteria, as codified by A Bradford Hill³⁴ or the 1964 Surgeon General's Advisory Committee, but epidemiologists today would do well to recognize the element of human judgment involved in their use.³⁵

Conflict of interest: None declared.

References

- Campbell AM. The effect of excessive cigarette smoking on maternal health. *Am J Obstet Gynecol* 1936;31:502
- Lilienfeld DE. Abe and Yak: the interactions of Abraham M. Lilienfeld and Jacob Yerushalmy in the development of modern epidemiology (1945-1973). *Epidemiology* 2007;18:507-14.
- van den Berg B, Christianson R, Oechsli F. The California child health and development studies of the School of Public Health, University of California at Berkeley. *Pediatr Perinat Epidemiol* 1988;2:265-82.
- Yerushalmy J, Hilleboe HE. Fat in the diet and mortality from heart disease. *NY State J Med* 1957;57:2343-52.
- Yerushalmy J, Palmer CE. On the methodology of investigations of etiologic factors in chronic diseases. *J Chronic Dis* 1959;10:27-40.
- Berkson J. Smoking and cancer of the lung. *Proc Staff Meetings Mayo Clin* 1960;35:367-85.
- Cornfield J, Haenszel W. Some aspects of retrospective studies. *J Chronic Dis* 1960;11:523-34.
- Yerushalmy J. Statistical considerations and evaluation of epidemiological evidence. In: James G, Rosenthal T (eds). *Tobacco and Health*. Springfield, IL: Charles C. Thomas, 1962.
- Lilienfeld AM. On the methodology of investigations of etiologic factors in chronic diseases—some comments. *J Chronic Dis* 1959;10:41-46.
- Simpson WJ. A preliminary report on cigarette smoking and the incidence of prematurity. *Am J Obstet Gynecol* 1957;73:807-15.
- U.S. Department of Health, Education, and Welfare. *Smoking and Health. Report of the Advisory Committee to the Surgeon General of the Public Health*. DHEW Publication No. 1103. Washington, DC: Communicable Disease Center, 1964.
- Yerushalmy J. Mother's cigarette smoking and survival of infant. *Am J Obstet Gynecol* 1964;88:505-18.
- Smoking during pregnancy. *Br Med J* 1968;4:339-40.
- Russell CS, Taylor R, Law CE. Smoking in pregnancy, maternal blood pressure, pregnancy outcome, baby weight and growth, and other related factors. A prospective study. *Br J Prev Soc Med* 1968;22:119-26.
- Schmeck HM. Surgeon General urges battle to dissuade women smokers. *New York Times*, 12 January 1971.
- US Department of Health, Education, and Welfare. *The Health Consequences of Smoking A Report of the Surgeon General: 1971*. DHEW Publication No. (HSM) 71-7513. Washington, DC: U.S. Department of Health, Education, and Welfare, 1971.
- US Agency seeks to curb smoking during pregnancy. *Louisville Times*, 30 July 1971.
- Tobacco Institute. Smoking and Pregnancy: Is the Question Answered? 1971. <http://legacy.library.ucsf.edu/tid/fbj62f00> (31 July 2014, date last accessed)
- Mothers needn't worry, smoking little risk to baby. *Boston Record-American*, 31 July 1971.
- Yerushalmy J. Infants with low birth weight born before their mothers started to smoke cigarettes. *Am J Obstet Gynecol* 1972;112:277-84.
- In defense of smoking moms. *Family Health*. May 1972: 11-12.
- U.S. Department of Health, Education, and Welfare. *The Health Consequences of Smoking. A Report of the Surgeon General, 1973*. DHEW Publication No. (HSM) 73-8704. Washington, DC: U.S. Department of Health, Education, and Welfare, 1973.
- The cigarette controversy. *Congressional Record. 93rd Congress, 1st Session. February 7 1973*;119:1-5.
- U.S. Department of Health and Human Services. *The Health Consequences of Smoking for Women. A Report of the Surgeon General*. Washington, DC: U.S. Department of Health and Human Services, 1980.
- Tobacco Institute. *Synopses on Women and Smoking*. 1980. <http://legacy.library.ucsf.edu/tid/how92f00/pdf> (31 July 2014, date last accessed)
- Jones KL, Smith DW. Recognition of the fetal alcohol syndrome in early infancy. *Lancet* 1973;2:999-1001.
- The Council for Tobacco Research. *Log: Special Projects*. 1992. <http://legacy.library.ucsf.edu/tid/faz36b00> (31 July 2014, date last accessed)
- Letter from Arthur J Stevens to Alexander Spears. 1972. <http://legacy.library.ucsf.edu/tid/xwv58c00> (31 July 2014, date last accessed)
- Statement of Richard J Hickey, 2 March 1982. In: Comprehensive Smoking Prevention Education Act of 1981. *Hearing before the Committee on Labor and Human Resources*. Washington, DC: US Government Printing Office, 1982.
- Parascandola M. Skepticism, statistical methods, and the cigarette: a historical analysis of a methodological debate. *Perspect Biol Med* 2004;47:244-61.
- Lilienfeld AM. Emotional and other selected characteristics of cigarette smokers and nonsmokers as related to epidemiological studies of lung cancer and other diseases. *J Natl Cancer Inst* 1959;22:259-82.
- Heath CW. Differences between smokers and nonsmokers. *AMA Arch Intern Med* 1958;101:377-88.
- Lilienfeld DE. Abe and Yak: the interactions of Abraham M. Lilienfeld and Jacob Yerushalmy in the development of modern epidemiology (1945-1973). *Epidemiology* 2007;18:507-14.
- Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.
- Parascandola M, Weed DL, Dasgupta A. Two Surgeon General's reports on smoking and cancer: a historical investigation of the practice of causal inference. *Emerg Themes Epidemiol* 2006;3:1.