

American Journal of Evaluation

<http://aje.sagepub.com/>

Realistic Evaluation Bloodlines

Ray Pawson and Nick Tilley

American Journal of Evaluation 2001 22: 317

DOI: 10.1177/109821400102200305

The online version of this article can be found at:

<http://aje.sagepub.com/content/22/3/317>

Published by:



<http://www.sagepublications.com>

On behalf of:



AMERICAN
EVALUATION
ASSOCIATION

American Evaluation Association

Additional services and information for *American Journal of Evaluation* can be found at:

Email Alerts: <http://aje.sagepub.com/cgi/alerts>

Subscriptions: <http://aje.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://aje.sagepub.com/content/22/3/317.refs.html>

>> [Version of Record](#) - Sep 1, 2001

[What is This?](#)

Realistic Evaluation Bloodlines

RAY PAWSON AND NICK TILLEY

INTRODUCTION

In this paper, we take the reader on a brief tour of six social scientific inquiries, old and new. They range from major pieces of work to quite minor ones. They employ contrasting research strategies and incorporate different value positions. They draw upon data and observations extracted from around the globe. Some of them cross-refer to each other and some do not. Not many of them can be thought of as “evaluations.” Despite all this, we feel they tell a realistic tale about how evaluation should be conducted and what it can achieve. The payoff from our whistle-stop trip comes at the end, where we spell out six evaluation lessons that can be discerned from what we find along the way.

As a final word of preface, we draw upon the sentiments of some famous Shakespearean opening-lines in which Chorus notes the folly of trying to “turn th’ accomplishments of many years into an hourglass.” And so in the same spirit we, “prologue-like, your humble patience pray—Gently to hear, kindly to judge, our play.”

BLOODLINES

Episode One: Social Policy, London School of Economics, 1970

Surgery often requires transfusion. Bad blood is dangerous for transfusion purposes. An unintended, iatrogenic consequence of using contaminated blood for transfusion may be the death rather than well-being of the patient. Richard Titmuss’ (1970) pioneering study assessed the then relative merits of differing ways of acquiring and distributing blood. The chief point of comparison for Titmuss was between blood donation where there is a market and blood donation where there is none. His normative preferences are never hidden and he finds strongly in favor of market-free donation. Though he refers to arrangements in a range of countries, Titmuss’ major focus was on the United States and Britain. In the U.K. then, as now, all blood was obtained by voluntary donation. In the U.S. there was a private market in blood.

Ray Pawson • ESRC Centre for Evidence Based Policy, Queen Mary University of London, Mile End Road, London E1 4NS UK; Tel: +44-0207-882-7657; Fax: +44-0207-882-7641; E-mail: Ray.Pawson@EvidenceNetwork.org.

American Journal of Evaluation, Vol. 22, No. 3, 2001, pp. 317–324. All rights of reproduction in any form reserved.
ISSN: 1098-2140 Copyright © 2002 by American Evaluation Association.

Titmuss draws together a wide range of information relating to blood collection and distribution. He critically reviews existing studies, assembles those administrative data that are collected, and engages in a modest amount of primary research to test the theory. Though imperfect in many respects, these studies consistently corroborated his normative theory. In the end, Titmuss is confident in his conclusions and in the significance of them. In a thundering rhetorical flourish he concludes:

... the commercialization of blood and donor relationships represses the expression of altruism, erodes the sense of community, lowers scientific standards, limits both personal and professional freedoms, sanctions the making of profits in hospitals and clinical laboratories, legalizes hostility between doctor and patient, subjects critical areas of medicine to the laws of the marketplace, places immense costs on those least able to bear them—the poor, the sick, and the inept—increases the danger of unethical behavior in various sectors of medical science and practice, and results in situations in which proportionately more blood is supplied by the poor, the unskilled (and) the unemployed. . . Redistribution . . . of blood and blood products from the poor to the rich appears to be one of the dominant effects of the American blood banking systems.

Moreover, on four testable nonethical criteria, the commercialized blood market is bad. . . (I)t is highly wasteful of blood; shortages, chronic and acute, characterize the demand and supply position. . . It is administratively inefficient and results in more bureaucratization. . . In terms of price per unit of blood to the patient (or consumer) it is. . . five to fifteen times more costly than the voluntary system in Britain. And, finally, in terms of quality, commercial markets are much more likely to distribute contaminated blood; the risks for the patient of disease and death are substantially greater. Freedom from disability is inseparable from altruism (Titmuss, 1970, p. 277).

Episode Two: Sociology, Princeton, 1994

Viviviana Zelizer, social theorist and historical sociologist, enters the domain of economics with iconoclastic results. She opines of the prevailing wisdom that “there is no question about the power of money to transform nonpecuniary values, whereas reciprocal transformation of money by values or social relationships is seldom conceptualized or explicitly rejected” (Zelizer, 1994, p. 2). Bit firmly between teeth, she proceeds to make a case against the notion that money always corrupts and commodifies. Her method takes her on a grand historical tour of the roles played by money in consumer life, welfare, and culture in nineteenth and early twentieth century America. In vignette after vignette, she demonstrates forces of acquiescence *and* resistance in everyday financial practices:

Consider, for instance, how we distinguish a lottery winning from an ordinary paycheck, or from an inheritance. A thousand dollars won in the stock market do not “add up” in the same way as \$1,000 stolen from a bank, or \$1,000 borrowed from a friend. A wage earner’s first paycheck is not the exact equivalent of the fiftieth or even the second. The money we obtain as compensation for an accident is quite different from our royalties for a book. And royalties gained from a murderer’s memoirs fall into a separate moral category from royalties earned by a scientific text (Zelizer, 1994, p. 5).

Her particular interest is in the process of “earmarking.” Much as people use different idioms, dialects, and accents in different social contexts, so pin money, blood money, paychecks, poor relief, and “other currencies” are set aside and valued in quite different ways. Zelizer retells the study of “contested earmarking” between charity workers and charity

recipients in the tenement district in the West Side of New York in the early 1900s. Much to the incomprehension of their middle-class benefactors, many immigrant families set aside a significant proportion of their meager poor relief in the form of “death money” to be used for the extravagant funerals of loved ones (neighbors would talk if there wasn’t a “fine layout”). The battle between the “sacred gift” and the “sacrilegious extravagance” lasted for many a year until welfare reform and cultural change saw it to a draw.

Episode Three: Medicine and Pathology Laboratory, Minnesota, 1998

Bad blood is not all the same. Hematological studies now identify a whole range of viruses, antigens, and infectious diseases. Accordingly, the comparison between altruistic donation and monetary inducement has by now been deepened clinically and widened by comparative study “in many countries, over many decades.” Eastlund’s (1998) meta-analysis, of which we provide a fragment in Table 1, captures a remorseless pattern.

This table shows that across twenty studies, collectively covering as many as 11 disease “markers,” volunteers were found consistently to outperform paid donors, and by very large margins. In the “best case” for paid donors (Okochi & Murakami, 1968), paid donors showed markers for HbsAg at close to twice the rate as that for volunteers (2.2% vs. 1.2%). At the other extreme, Walsh et al. found paid donors showing markers for nonspecific hepatitis at over 50 times that for volunteers. Indeed, 0% of volunteers in this study were found to have markers as against 51% of those paid.

TABLE 1.
Disease Markers in Paid and Volunteer Blood Donors

Reference	Marker	(Percentage of donors with marker)	
		Paid (%)	Volunteer (%)
Allen et al.	Nonspecific Hepatitis	4.1	0.7
Seeff et al.	Nonspecific Hepatitis	15.5	2.6
Walsh et al.	Nonspecific Hepatitis	51	0
Koretz et al.	HBV	41	0
Allen et al.	HBV	36.7	10
Hollinger et al.	NANBH	28.1	7
Aach et al.	NANBH	34.6	6.7
Seeff et al.	NANBH	17.3	6.9
Hayashi et al.	HCV	3.3	0.4
Okochi et al.	HBsAg	2.2	1.2
Goesser et al.	HBsAg	1.57	0.11
Prince et al.	Anti-HBs	38.5	4.9
Szmunness et al.	Anti-HBs	20	6
Dawson et al.	Anti-HCV	10.08	0.36
Jna et al.	Anti-HCV	13.91	0.25
Hughes et al.	Anti-HIV	0.04	0.015
Abbott	Anti-HIV	0.133	0.014
Canavaggio et al.	Anti-HTVL	0.3	0.025
Dile et al.	Anti-GBV-CE2	34	3.8
Abbot et al.	HGV RNA	13–15	1

Source: Derived from Eastlund 1998

Episode Four: Sociology and Anthropology, Athens, Ohio, 1999

Our next inquiry is a finely textured case study of one particular group of donors. The “blood industry” in the U.S. is bifurcated, with a nonprofit (Red Cross) system existing alongside the commercial sector. The commercial system itself is also split, using two contrasting collection points—namely, “downtown” centers drawing from low-income donors and “college” centers located near large universities (Anderson, Newell, & Kilcoyne, et al., 1999).

The former may be considered emblematic of Titmuss’ vision of the bloodsucking of the poor by the rich—but this is hardly a good description of the latter. The Ohio researchers give some vital clues to the nature of the cultural expectations about donation in such circumstances. Their survey demonstrates that student donors are in fact slightly better off than the undergraduate norm. And this touch of affluence brings with it a further set of characteristics that color the plasma collection operation. Students donors, it seems, are more likely to smoke and have higher levels of alcohol consumption than their fellow scholars; they are, in short, more likely to be “party animals” (Anderson et al., 1999, p. 150).

While none of this conjures up a particularly inspiring vision, it is clear that we are dealing with neither “altruism” nor “economic dependency.” And while the Ohio team do not have the range of hematological tests at their disposal, there seems no reason to suppose that bad blood flows from the arrangement. Indeed, these students tend to quit selling plasma unless they are in the best of health (Anderson et al., 1999, p. 155). So how is the incentive working for this group? What are the cultural norms involved in seeking and accepting payment? Anderson and colleagues discover the unpleasant little chore is performed because it produces a “treat” that falls outside normal budgetary constraints. In short, the students “earmark” their fee for downright frivolous activity:

I kind of considered it just like getting 20 bucks from your grandmother. Even if you don’t have money, it’s 20 free dollars. You’re going to go out and blow it on something. So, that’s how I always—I never really needed it, essentially, but it was always useful (Anderson et al., 1999, p. 149).

Episode Five: Public Health and Law, Columbia and New York Universities, 1999

So far, our story has shown sturdy support for a blood collection regime based on altruistic donation plus some modest evidence for a place for earmarked payments. What other factors might influence the design of a system? Here we come to the bugbear of all policymaking. The issue is summed up for us in the patricianly thoughts of former British prime-minister, Harold Macmillan, who when asked to recall the greatest difficulty he faced in governing the country, replied “events, dear boy, events.”

Many significant events have occurred since Titmuss’ time. The medical, scientific, cultural, and social contexts for collection and distribution of blood has been transformed. AIDS has arrived. Many people have died because of bad blood inadvertently given and transfused by altruistic donors. Gay rights have progressed substantially and discriminatory policies and practices have become anathema. For many diseases, blood can be and is routinely tested in ways that were not possible in the 1960s. Advanced methods of purification have developed. There is also much greater sensitivity to and aversion to risk—indeed, we are deemed to live in a “risk society.” We thus turn next to a reflective study that places

such changes in the context of Titmuss' grand thesis on altruism. In Bayer and Feldman's (1999) view:

The aura that has come to surround volunteer blood donors and altruistic blood donation has exacted a price that only became apparent in the context of the AIDS epidemic. Because donors were viewed as selfless, and because the process of donation was viewed as an expression of solidarity, it was politically and ethically difficult to develop policies that distinguished among potential donors. Blood authorities could not simply exclude those who might pose a risk because of their behavior or because they came from nations or groups thought to present increased risk to the blood supply. Indeed, some of the most contentious encounters in the nations as diverse as the United States, Australia, and Denmark (in terms of the status of gay men) centered on the potential benefits from and consequences of efforts to exclude gay blood donors, who viewed such actions as a manifestation of homophobia and a threat to the goal of social equality. . . . In the aftermath of the AIDS epidemic, the mythic equivalence of the voluntary donor and the safe donor has been shattered (Bayer & Feldman, 1999, p. 8).

And what of technology? The advent of tests for contamination and techniques of purification can dispose of much of the potential bad in blood. Do they also dispatch Titmuss' theory to antiquity? Inevitably costs are high and economics intercedes in the quest for purity. Perhaps more importantly, technology is imperfect and Bayer and Feldman show how a little learning can be a dangerous thing. In this respect, they provide evidence for the continued relevance of Titmuss' model when they tell us that:

Although Vietnamese blood is screened for HIV antibody, such testing can fail to detect the presence of infection during the first months after a donor contracts HIV. Where the incidence of new infections is high this can pose a serious threat to blood safety. Moreover, because the majority of blood is collected from professional blood sellers in urban areas, where rates of HIV are higher than in the countryside, blood shipped from the city is introducing HIV into areas that previously had low rates of infection. A similar pattern has emerged in China, making the blood supply a major source for the spread of AIDS (Bayer & Feldman, 1999, p. 15).

Episode Six: Economics, Waterloo, Canada, 1996

For our final case study we head back in time, leave behind the blood feuds, and indeed step outside the concerns of public policy altogether. We turn to the seemingly technical issue of "nonresponse" in social surveys. Objectivity is the single-minded concern of practitioners here. Such a goal is often dealt a blow at the very first hurdle, with surveys often facing miserable response rates of less than a third. Offering "incentives" to respondents has long been mooted as a possible solution to the problem, and this leads us back to a more familiar debate about whether "egoism" or "altruism" should be cultivated as the basis for response rate bliss.

Warriner et al.'s 1996 study is of interest because it puts this old puzzle to a beguiling empirical test. Respondents were allocated to different "inducement conditions" for completing and returning a survey—some were offered an entry into a lottery draw, some were promised a donation on their behalf to charity, and some were confronted with cash. Each "treatment" was also varied in size so that no less than twenty different conditions were assigned throughout the sample.

And the results? The main title of the paper is "Charities, No; Lotteries, No; Cash, Yes."

This together with the key finding that the cash incentives produced significantly better response rates across all classes of the population might seem to suggest that, even in Canada, altruism lies dormant. In fact, the authors do not draw such a straightforward solution for their real interest lies in some of the minutiae in the data, "... incentives in the amount of \$2 and \$5 give appreciable increases to the response rate, but the increment from using \$10 in place of \$5 is negligible" (Warriner et al., 1996, p. 550). The authors' interpretation is that the modest cash incentive works because respondents view it as a bit-of-a-treat-for-a-bit-of-a-chore. They perceive the hand of "reciprocity" in the act of response, lying in the "middle conceptual ground between more subtle concepts of helping behavior on the one hand or a nakedly economic self-interest interpretation on the other" (Warriner et al., 1996, p. 559). Earmarking, it seems, has surfaced again.

IMPLICATIONS

Evaluation research is cursed with "short-termism." Programs are dispatched to meet pressing dilemmas, evaluations are let on a piecemeal basis, methods are chosen to pragmatic ends, and findings lean towards parochial concerns. Our hope, possibly against hope, is for a future evaluation culture that is more painstaking and for an evidence base that is more cumulative. To this end, we come to the concluding episode of the bloodlines, in which we draw out some implications as a series of six maxims for evaluations yet to come.

1. *Always speak of evaluations in the plural.* Here we echo a point made forcefully by Mark, Henry, and Julnes (2000, p. 72), who argue that only a tailored portfolio of studies can cope with the profession's multiple goals of evaluating "merit," "worth," "improvement," and "compliance" in an initiative. We have made a related point ourselves (Pawson & Tilley, 1997, p. 147) in eschewing the "one-off" approach to evaluation and in demonstrating the cumulative power of an iterative series of inquiries following the fortunes of the same policy line. But now we want to go further. Both of these suggestions assume a logical guiding hand shoving evaluation along steadily to a rational future. It may be better, as in our current example, to rely on a touch more serendipity and to raise the principle of always scouting widely for strong shoulders upon which to stand. The little collection brought together here tells an evaluation story, but one that can and could only be culled post hoc.
2. *Be unafraid to ask big questions of small interventions and to use small interventions to test big theories.* Titmuss began with a specific problem to do with acquiring and distributing blood for transfusion purposes. He addressed it using rather grand classic theory about gifts and about forms of exchange. He drew from his findings some ambitious conclusions about principles of social policy. Titmuss' start and end points were theory. His data collection spoke to theory. Potentially fecund policy-relevant tested theory was the pay-off, so that modern commentators like Feldman and Bayer perceive their work as revision and refinement of the body of knowledge. Likewise, Zelizer finds the courage of her conviction to oppose theories of the cash nexus (monetarist or marxist) in small domestic and community episodes. And we bring further refinement to her theory here with the discovery of "earmarking" at work in blood donation and survey returns.

3. *Use multiple methods and multiple data sources in the light of opportunity and need.* Having been conducted decades before the advent of the evaluation industry and the rhetoric of evidence-based policy and practice, Titmuss' study has not often been considered an "evaluation" but by golly it is, and an exemplary one to boot. What Titmuss does is to compare the types of contexts for blood donation furnished by market and nonmarket arrangements. He makes sense of the differing mechanisms that are triggered in each, by showing the different forms of reasoning and resources that are facilitated. And he identifies contrasting outcomes that are to be expected under the two regimes. In the language of realism, he constructs middle range context, mechanism, outcome configurations (CMOCs). He then tests the theories, by critically and opportunistically gathering together all the information that speak to them, and supplementing this where necessary (and as resources permit) with primary research. In practice, this meant a survey of 3,800 British donors that Titmuss consistently refers to as a "pilot." Titmuss spent only £3,250 (roughly £31,000 or \$43,000 today)—rather good value for money by current standards. The further studies in our sequence add unstructured interviews, meta-analysis, documentary analysis, and logistic regression to the list, making the point, yet again, that the only methodological gold standard is pluralism.
4. *Figure out which mechanisms are relevant to produce optimum outcomes by context.* Contexts have multiple layers. A given policy or measure is liable to activate multiple mechanisms with divergent outcomes. Similar outcomes may in principle be generated through varying mechanisms. So, we need to ask what balance of benefit versus harm is best produced through elements of trust, honesty, altruism, payment, laboratory tests, and processes of purification, in what context. If every available check, every test and every available means of purification is used, the price of blood is increased but the potential harm from its transfusion minimized. In an individualistic, rich, risk-hypersensitive, litigious society the balance may be different from that in a collectivist, poorer, and less litigious one. In the latter, altruistic donation may make more sense, with fewer tests, providing more blood, to be allocated on the basis of need.
5. *Never expect to know "what works," just keep trying to find out.* Titmuss may or may not have believed he was producing a universal truth. He was not doing so. Contexts change. New phenomena with new attributes emerge, altering the conditions for and consequences of action. Titmuss wrote before the AIDS epidemic. Initiatives themselves transform systems—the development of bad blood-tests changes the odds of identifying contaminated blood from otherwise risky donors. Moreover, circumstances vary. Consider our tale of survey response incentives. Before we advise all surveys companies and research institutes to include the newly minted dollar in the questionnaire schedule, we should pause for thought about the specifics of context. Warriner et al. do not mention the "topic" of the survey under the incentives experiment. Might not this change the potential respondents' reactions to incentives of varying levels? The tariff for obtaining our views on shampoo preferences might be of a different order from that for our opinion on compulsory bloodletting. Moreover, the incentives effects for these issues might vary for the hairless and hemophiliacs compared to the hirsute and healthy.
6. *Direct meta-analytic inquiries at common policy mechanisms.* The notion that "incentives" crop up with the aim of increasing blood donation and boosting survey

response comes as no surprise. It is less obvious that an “earmarking” response should follow in some quarters. This gives a clue as to where the real learning lies. Policies are contagious. News of the success of “naming and shaming” in *A*, is followed by the launch of “naming and shaming” in *B*, *C* and *D*. “Zero tolerance” of *E* begets zero tolerance of *F*, *G*, and *H*. Successful “multiagency working” in *I* has agents multiplying in *J*, *K*, *L*. The fact that the same (old) ideas tend to be tried over and again in quite different contexts is an absolute blessing to the evaluator. In stretching initiatives to and beyond their limit, policymakers provide evaluators with a natural canvass with which to discover when and why policy really works. Meta-analysis has tended to locate itself within tightly restricted policy domains and concentrated on evaluating rival means of treating the same end. We believe that more is to be learned by analyzing the same means applied to different ends (Pawson, 2002).

And, finally, a summary surmise on the consequences of all this for the future of the evaluation profession. We have argued that good evaluation is good social science. For us, this embraces the gallant aims of precision in articulation of theory, rigor in empirical testing, confederation in lines of inquiry, and cumulation in the body of findings. The “realist movement,” of which we are a part, is often considered the brash upstart of the evaluation schools. In fact, it depends on these rather venerable ideas. The future, for us, thus lies in keeping faith with some of the grand old principles of social science and in not forgetting the hard-won lessons of the old studies.

REFERENCES

- Anderson, L., Newell, K., & Kilcoyne, P. (1999). Selling blood: Characteristics and motivations of student plasma donors. *Sociological Spectrum*, 19, 137–162.
- Bayer, R. & Feldman, E. (1999). Introduction: Understanding the blood feuds. In E. Feldman & R. Bayer (Eds.), *Blood feuds*. New York, NY: Oxford University Press.
- Eastlund, T. (1998). Monetary blood donation incentives and the risk of transfusion-transmitted infection. *Transfusion*, 38, 875–882.
- Mark, M., Henry, G., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.
- Okochi, K. & Murakami, S. (1968). Observations on Antigen (in Japanese) *Vox Sang*, 15, 374–385.
- Pawson, R. (2002). Evidence-based policy: In search of a method. *Evaluation* 8,(2).
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Titmuss, R. (1970). *The gift relationship: From human blood to social policy*. London: Allen & Unwin.
- Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & Mespurren, K. (1996). Charities, no; lotteries, no; cash, yes: Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly*, 60, 542–562.
- Zelizer, V. (1994). *The social meaning of money*. New York, NY: Basic Books.