# Lecture # 13 – Evaluating Public R&D

## I. Challenges of Evaluating Science Policy

- While policymakers assume that investments in science will increase economic growth, success is not uniform
  - For example, Japan's investments were not sufficient to prevent the economic decline of the 2000s
  - Understanding why results differ across countries is important, but challenging
- Lane discusses challenges for evaluating science policy
  - Relationship is non-linear
    - Outcomes vary by discipline
    - Demand and supply sides "intertwined"
  - Time lags are long
    - Simply looking at the jobs directly generated (e.g. research positions), as is done in the example from the Information Technology and Innovation Foundation's report on the science stimulus, is not sufficient
  - Many different units of analysis
    - Individual researchers and project teams
    - Research organizations (e.g. firms, universities)
    - Political systems
  - As a result, analysis of science policy is interdisciplinary
    - Science may have benefits that are not easily monetized, such as safety, security, environment
  - U.S. scientific data infrastructure oriented towards administration, rather than empirical analysis
    - At the end of class, we'll discuss whether management of science could be changed to better accommodate evaluation.

## II. Methods for Evaluating R&D

- Background on NIST report on evaluating R&D
  - Put together by Advanced Technology Program (ATP) at National Institute of Standards and Technology (NIST)
    - ATP has done several evaluations of publicly funded R&D
      - ATP is part of NIST, which is part of the Dept. of Commerce
      - Authorized in 1988
    - Goal of this report was to provide guidelines for doing these evaluations
- Evaluation best practices
  - Study scope and rigor

        - Will it be comprehensive, looking at all stages of a program, or focus on certain program aspects
            - Depends on budget for evaluation
        - Will it be statistically rigorous or look for sufficient evidence available in a limited amount of time?
    - What is program success?
        - Is the goal to be cost effective, to meet a certain target, to have distributional effects…?
    - What is the counterfactual?
        - That is, what would have happened without the program in pace
        - Goal is to rule out competing explanations
        - Using a control group is one way to demonstrate effect
            - Can be difficult to find, however, in social sciences
            - Compare to medicine, where randomized experiments are used
        - When not available, expert judgment often used to establish a counterfactual
            - Less rigorous, but more feasible
    - Should evaluation be in house or from outside
        - Advantages of outside evaluation
            - Outside evaluators establish credibility
        - Disadvantages of outside evaluation
            - Takes time for outsiders to learn about programs
            - Outsiders may have their own agenda
                - Won't be credible if have established biases
            - Need to "translate" analysis into language suitable for stakeholders
            - Program participants may be reluctant to share data with outsiders
- The logic model
    - Program evaluation begins with a logic model
        - Presents a diagram of basic program elements
            - What does the program do?
            - How does it do it?
            - What are the intended consequences?
        - Shows linkages between components
        - Example in paper shows a generic logic model
            - Begins with a societal goal
                - Provides the impetus for a public program
            - Program is designed to meet the goal
                - Must obtain resources to carry out its mission
                    - These resources are inputs to the program
                        - Cost to the public
            - Outputs vary depending on the time frame
                - Consider a program to fund health research

                - Short term outcomes may be patents or publications
                - Long term outcomes may include development and use of new treatments
                - Even longer term may include increased life spans, reduced disease incidence, etc.
            - These impacts should be assessed against program costs

- Choice of evaluation method
    - There are several different evaluation methods commonly used
    - Below we discuss the strengths and weaknesses of each type
    - Choice depends on features such as the evaluation budget and the scope of the program
    - We'll leave econometric analysis for last, as Jaffe's paper discusses in more detail
- Surveys
    - Provide descriptive data on a program
    - Surveys can be open-ended or close-ended
        - Open-ended surveys need to be codes consistently
    - Choice of sample important
        - Is it representative?
    - More detailed data are possible (e.g. face to face interviews), but costly
    - Advantages
        - Cost-effective way to gather aggregate data
        - Can be used to establish counterfactuals if used early in program
        - Data collected can be inputs for other analyses
    - Disadvantages
        - Survey responses may be subjective
        - Will respondents be truthful?
            - Particularly an issue if need to recall past behavior
- Case study
    - "Descriptive case studies are in-depth investigations into a program" (p. 34)
    - Useful in exploratory phases of a program
    - Usually start with information about a program from direct observation
    - Advantages
        - Because written as a narrative, aimed at a broad, non-scientist audience
        - Provide richer detail than other methods
    - Disadvantage
        - Anecdotal evidence is less persuasive
        - Will the case be generalizable?
        - Is there selection bias?
            - That is, are cases more likely to look at successful than unsuccessful programs (my addition – not in article)

- Case study with economic estimation
    - Combines a descriptive case study with quantification of costs and benefits
    - The descriptive case study provides essential background on the program being analyzed
    - Can be prospective or retrospective
        - Retrospective uses past data to evaluate program
            - Thus, requires a longer history of data
        - Prospective makes projections of future effects
    - Challenges
        - Separating out roles of various contributors
            - That is, what can be attributed to the program, rather than other market forces?
            - Becomes more difficult the further upstream (e.g. closer to market) that the program is
        - Uncertainty and risk
            - Presenting quantitative results may make them look more certain than they really are
            - Adding probabilities to outcomes allows for risk assessment
            - If not possible, sensitivity analysis should at least be done
    - Advantages
        - Extends from start to finish of program
        - Provides quantitative estimates of results
            - Often more convincing than a descriptive case study
            - Allows for comparison (e.g. by comparing NPV or IRR)
    - Disadvantages
        - Can the dollar value of benefits be easily estimated?
            - Particularly difficult for non-market benefits, such as health or environment
                - E.g. what is the value of a life saved?
        - For projects not yet complete, may have negative IRR in short run
            - This may discourage key stakeholders
            - Could discuss Solyndra here
                - Was an unsuccessful solar panel manufacturer that had been supported by the government
                - However, if the goal was promoting solar PV, should expect some risk and some unsuccessful outcomes
        - Are results generalizable?
- Sociometric/social network analysis
    - Studying the structure of relationships
    - Uses both qualitative and quantitative methods
    - Goals:
        - To learn more about social networks and how they affect behavior
        - To learn about spheres of influence of scientists
        - To identify pathways of knowledge spillovers
        - To improve success of collaborative relationships

- How to identify networks
  - Patent/publication citations
    - Could also look for co-authorship
  - Ask participants with whom they share information
    - Then, ask the people mentioned the same question
  - Co-nomination analysis
    - Ask researchers in a given field to nominate others most similar to or most relevant for their own work
      - Assume links exist if people are co-nominated
- Advantages
  - Show connections in a way that other analyses cannot
- Disadvantages
  - Unfamiliar to many key stakeholders
  - How to interpret qualitative measures?
    - Are they meaningful to ultimate program goals?

- Bibliometrics
  - Use of publication and patent data
    - Both are key outputs of scientific programs
  - Ways to use these data
    - Simple counts
      - Can normalize by research costs
      - Can adjust for quality of journals
      - As discussed earlier in the class, not all patents or publications are equal
    - Citation analysis
      - Can be a measure of quality
      - Can indicate knowledge flows
        - Program may want to know if it is influencing other sectors of the economy
    - Content analysis
      - Examples include searching for frequency of key words
  - Advantages
    - Widely applicable
    - Data readily available
      - Don't need to collect data from program participants, because other databases exist
  - Disadvantages
    - Ignores other outputs
      - May miss long-term outcomes
        - E.g. do publications about a possible new medicine eventually lead to treatment?
    - Need to adjust for quality
    - Propensity to patent and publish varies across fields
- Historical tracing
  - Like a descriptive cast study
    - Typically involves interviews and qualitative methods

    - Traces a series of interrelated developments chronologically
    - Forward tracing begins with the research if interest and traces the evolution of related events from that point forward
    - Backward tracing starts with an outcome of interest and traces backward to find the key events that led to the outcome
        - More appropriate when the outcome is the central focus
    - Advantages
        - Produces credible outcomes
    - Disadvantages
        - Chain of events can be complex
            - Many players may be involved
        - Dead ends can slow the research
- Expert judgment
    - Ask experts for opinions on quality and effectiveness of a research program
        - Experts must know the subject
        - Must have no conflicts of interest
    - Can also be used to make predictions about future outcomes (expert elicitation) – not in article
    - Advantages
        - Practical
            - Quick and straightforward
        - Can lead to exchange of ideas with experts
    - Disadvantages
        - Is it accurate?
        - Do experts have biases?
- Econometric methods
    - Statistical analysis of functional relationships between economic and social phenomena
    - Possible uses
        - Test hypotheses
            - Requires a theoretical model to motivate the hypotheses
            - Suggests establishing a causal relationship
        - Establish correlations
            - No attempt to establish causality
            - Calculates correlations conditional on other variables
                - Thus, more useful than simple pairwise correlations
        - Estimate a production function
            - Relates inputs and outputs
        - Develop a macroeconomic model for forecasting
    - Advantages
        - Can determine causal relationships
    - Disadvantages
        - May be more difficult for a non-technical audience to understand
        - Requires sufficient data

## III. Econometric Analysis of R&D programs

- Jaffe's paper discusses the research challenges of using econometric techniques to evaluate public R&D programs

The major issue is the selection bias problem

  - Consider a standard research program
    - Successful applicants are chosen after peer review
      - Typically merit review, but other criteria such as diversity of gender, race, or geography may also matter
    - Receipt of funds is not conditional on specific output
  - Standard regression model:
    - $Y_{i,t} = bD_i + lX_{i,t} + a_i + m_t + w_{i,t} + e_{i,t}$
      - $Y_{i,t}$ is research output (e.g. publications)
        - Note, for now, ignoring problem of lags
      - $D$ is a dummy variable = 1 if individual *i* receives a grant
        - $b$ is the effect of receiving a grant
        - This is our coefficient of interest
          - We estimate the average effect of all those receiving funding: $E(b|D)$
      - Other controls include firm size, age, demographics, etc.
      - Finally, include fixed effects
        - $a_i$ represents applicant effects
        - $w_{i,t}$ represent period and applicant specific variation in research productivity *that is not observed by the researcher*.
          - However, it is observed by the granting agency
          - What might this represent
            - Skill or reputation of researcher
            - Quality of the proposal
      - It is the unobserved $w_{i,t}$ that cause selection bias
        - The best projects (e.g. highest $w_{i,t}$) are chosen
        - Leads to omitted variable bias, because $E(a+w \mid D) > 0$
  - Possible solutions for selection bias
    - Regression with controls
      - Arora and Gambardella (1998) consider the effect of NSF funding on economist publications
        - Use impact-weighted publications in a 5-year window after funding as output measure
        - Include impact-weighted publications in the 5 years prior as a control
          - Provides a measure of existing ability
    - Matched samples of treated and untreated entities
      - The challenge is whether you can select comparable groups using observable data only
    - Fixed effects (a/k/a "differences in differences)

      - Using panel data allows the researcher to eliminate time-invariant effects
        - Doesn't provide information on the quality of the specific proposal receiving the award
    - Model the selection process/instrumental variables
      - Find a variable that affects the probability of selection but does not affect performance
        - Often depends on institutional knowledge of the process

How could R&D programs be designed to facilitate evaluation?

- Randomization
  - Identify a group of potential awardees
    - Thus, don't have to just give money to anyone
      - Could even have multiple groups: high priority, marginal, and rejects
        - Could have higher probability of funding for high priority
        - Then, simply need to evaluate each group separately
    - Doesn't bias effect on the treated group
      - However, couldn't generalize the results to those whose proposals were screened out as inferior
  - Randomly give wards within this group
    - Thus, the probability of getting an award is the same for all members in the group
  - Those that don't get chosen are the control group
  - Are there concerns for implementation**?**
    - Some high quality proposals may not be funded
      - Perhaps at the expense of lesser proposal being funded
- Regression discontinuity
  - Use discontinuity at thresholds of acceptance to identify similar proposals
    - Rank all proposals
    - Select a threshold
      - All above get funding, all below do not
    - Can now estimate our equation using the ranking as a control variable
    - Compares outcomes of proposals near the threshold
    - shows a break at the threshold
      - The performance of projects above the threshold is better than those below
      - The dummy variable controlling for the threshold represents this shift
  - Jacob and Lefgren (2011) provide an example of regression discontinuity

    - Consider the effect of NIH funding on publications and citations
        - Ask students what the selection bias would be
    - The NIH review process
        - Applications received at three times during the year
        - Non-blind peer review on five criteria
        - Each criteria given a score of 1-5
            - Average of these scores multiplied by 100 to give a priority score
            - Normalized for different research groups
        - Awards generally made based on priority scores
        - Funding depends on the agency's budget
            - Thus, the cutoff varies in different years
            - Figure normalizes scores so that 0 is the score of the last funded application (assuming none funded out of order)
            - Probability of eventual award shows if they reapply and receiving funding in a later competition
    - Because some funded out of order, not a perfect R&D
        - Instead, estimate an equation predicting funding as a function of the priority score and applicant characteristics
        - Final estimation uses this predicted probability of funding as a RHS variable to explore effect on publications
    - Find only a small impact of funding
        - Average grant is about $1.7 million
        - Leads to about 1.2 new publications (about a 7% increase)
        - Suggests researchers who don't get NIH funding can get funding elsewhere
            - Raises the possibility of crowding out
        - Also find that receiving an NIH award increases the amount of funding received in the next five years
            - The "Matthew Principle"–"success begets success"
- Can we measure "additionality"
    - That is, does public R&D support lead to new research, or replace work that would have been done anyway?
    - Selection bias an issue here as well
        - Firms more likely to do research are more likely to seek out support
    - Thus, Jaffe proposes a modified version of equation (1):
        - $Y_{i,t} = lX_{i,t} + + g^P P + g^G G + a_i + m_t + w_{i,t} + e_{i,t}$
        - $P_{i,t} = lZ_{i,t} + dG_{i,t} + a_i + m_t + w_{i,t} + e_{i,t}$
            - *P* represents private R&D

        - *G* represents government R&D
        - The second equation models the private R&D decision
            - *Z* are variables that affect private R&D, but not government R&D decisions
                - This is the challenge for the researcher
            - $g^G$ tells us the direct effect of government R&D on output
            - *d* tells us whether there is crowding out or crowding in
                - Crowding out: government R&D replaces private R&D
                - Crowding in: halo effect of receiving funding leads to more funding
                    - Receiving funding is a signal of quality
- Jaffe raises the question of whether evaluation should be built into the design of R&D programs
    - How might management of science be changed to make evaluation easier?
    - Should it be changed in such ways?
        - Currently, most evaluation is done after the fact
            - Building plans for evaluation into R&D programs could help solve selection bias
        - Randomization is one possibility.  Can students think of others?
        - Note that policy design for evaluation is not unusual in public policy
            - Randomization is particularly popular in developing countries
            - Has been used for school choice (e.g. lotteries)
            - Random selection