PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2010 - 31

The birth of the congressional clinic

Raphaël Godefroy

JEL Codes: H4, I1
Keywords: Health Policy, government policy, publicly-provided goods, medical research, legislative bargaining.

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE – INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE

The Birth of the Congressional Clinic*

Raphael Godefroy

Paris School of Economics

October 11, 2010

Abstract

This paper investigates the impact of mortality in the districts/states represented in key congressional groups (i.e. committees, subcommittees, and parties) on the public investment in medical research in the US. I focus on National Institutes of Health (NIH) R01 grants awarded between 1985-2002. Exploiting the recomposition of any group after congressional elections, I estimate that the composition of the House Subcommittee on Labor, Health and Human Services, Education and Related Agencies (HouS), impacts the NIH budget: a 1% increase of life-years lost because of a disease in the districts represented in HouS increases the funds for clinical research on that disease by 1.2-3.2%. I also find that this impact results from the larger bargaining power of HouS or the House majority, or both groups, in the budget process. No group significantly impacts the allocation of funds for basic research, or the allocation of funds across states.

Keywords: Health Policy, Government Policy, Publicly-Provided Goods, Medical Research.
JEL Classification Number: H4, I1

1 Introduction

This paper studies the effect of the composition of major groups in Congress — i.e. committees, subcommittees, and parties — on the public budget for medical research in the US. The social cost of a disease, such as its impact on life-years lost or its mortality rate, varies across districts and across states,

*I am extremely grateful to Jay Bhattacharya for his generous support and advice on this project. I would also like to thank Romain Aeberhardt, Luc Behaghel, Tim Bresnahan, Caroline Hoxby, Matt Jackson, Seema Jayachandran, Ryan Lampe, Aprajit Mahajan, Neale Mahoney and Eduardo Perez for helpful comments. I also thank Dana Goldman, Margaret Blume Kohout, Daniella Meeker who helped me to retrieve part of the NIH data and to become more acquainted with the NIH grants.

and hence across congressional groups. In any given group, the social cost of a disease may also change over time, due to the possible recomposition of the group following congressional elections, which occur every two years. I exploit these changes to estimate the effect of the change of life-years lost due to a disease on the change of federal funds for research on that disease distributed through the National Institutes of Health (NIH) R01 grants. For this purpose, I construct a data set gathering information on NIH-funded projects for the period 1985-2002, combined with data on mortality across districts and states, and with data on the composition of relevant congressional groups. I find that the composition of one group, the House Subcommittee on Labor, Health and Human Services, Education and Related Agencies of the Appropriations Committee (HouS), has a distinct influence on the allocation of public funds for *clinical* research across diseases, where, by definition, clinical research, as opposed to *basic* research, encompasses any research project that involves human subjects, such as clinical trials. An increase of 1 percent of life-years lost because of a disease in the districts represented in HouS increases the funds for clinical research on that disease by 1.2-3.2 percent. On average, this effect corresponds to an increase of yearly funds for clinical research equal to around $3600 for every additional life-year lost in the population of the districts represented in HouS.[1]

The hypothesis that profit incentives impact private investment in R&D, and, ultimately, the nature of scientific discoveries, has opened a vast area of research (e.g. Drandakis and Phelps 1966, Samuelson 1965, Hayami and Ruttan 1970, Romer 1990, Grossman and Helpman 1991, Aghion and Howitt 1992 and Acemoglu 2002), and several empirical studies aim to distinguish this impact from that of the factors that may improve the relative productivity of R&D in some given field (see Section 2). Public funding of R&D has been much less studied than private investment, however, even though abundant normative work supports the value of that funding (e.g. Arrow 1962, Griliches, Klette and Møen 2000, and Murphy and Topel 2006).[2] Further, the amounts of private and public investments in medical R&D have been comparable for the last decades — for instance, the NIH budget and the private investment in pharmaceutical research amounted respectively to around 30 and 40 billion dollars in 2006 — and could thus have comparable consequences on medical innovations.[3] NIH R01 grants, which are the most common and best-documented grants, represent more than half of the NIH

[1] Unless mentioned otherwise, all the monetary amounts in this paper are in dollars of 2000.

[2] As stated by Griliches, Klette and Møen 2000: "There is [...] little controversy among economists about the desirability of governmental support to these activities."

[3] See "AAAS Guide to R&D Funding Data," available on http://www.aaas.org/spp/rd/guide.shtml.

budget and constitute a primary source of information on the distribution of federal funds for medical research, since they usually fund circumscribed projects that can be matched to a single disease or to a small set of diseases.[4] Focusing on these grants, the main contribution of this study is to raise a determinant of investment in medical R&D that does not stem from either profit incentives or from the relative productivity of medical R&D across diseases, under the assumption that these other factors are not correlated with changes in mortality within congressional groups caused by electoral outcomes.

How do the results of the effect of the composition of HouS on clinical research inform the functioning of Congress? If the parameters that define the utility a congressman derives from medical research are uncorrelated with his or her membership status in HouS — in short, if HouS members are not preference outliers — these results show that HouS members have an effective larger bargaining power on the NIH appropriations than the rest of the House, and than the other groups I consider. Since that assumption could be questioned, I perform an instrumental variable analysis and other tests to estimate the actual power of HouS members on the allocation of NIH funds. Due to the correlation between mortality conditions in HouS districts and the districts represented in the House majority, these estimations do not permit distinguishing the respective bargaining power of the representatives of these two groups. However, these estimations and additional estimations of the influence of the House majority party do indicate that the members of one or both of these groups have an advantage in the bargaining process that decides on the allocation of funds across diseases, and use this power to favor clinical research on diseases that affect more their constituents. No other group has any significant power on that allocation.

Congressmen could use their bargaining power to meet local objectives other than supporting research on their constituents' diseases: they may want to supply funds to the research institutes located in their jurisdictions. I investigate that point and find no robust impact of the number of representatives of a state represented in HouS, or in any other group, on the sum of funds received by the research institutions located in that state. Earmarking of federal funds for research to specific locations, although raised in numerous anecdotes and empirical studies (see Cohen and Noll 1991 and

[4]This budget, dedicated to medical research, consumes around half of the public budget for non-defense research, much more than the National Science Foundation (NSF). The NIH is the main agency in charge of distributing federal funds to research projects on biomedical topics. Other agencies may also award research grants on similar topics, such as the NSF and the Department of Health and Human Services (HHS). The HHS supports mainly health services and training, and the NSF may support any area of scientific research, with a budget almost equal to a fifth of the NIH budget on average for the period studied here.

Section 2), has no substantial effect on the allocation of funds across states for the specific type of NIH grants I consider.

Since congressmen, with the President, have full institutional authority on the allocation of NIH funds, why don't they use their bargaining power to influence the allocation of funds across states, or the allocation of funds for basic research across diseases? To address these questions, I present some evidence suggesting that congressmen are concerned with the productivity of medical research, but lack information on it. With a simple model and anecdotal facts on the organization of the NIH, I argue that their concern for research productivity can lead congressmen to delegate their authority on the allocation of funds to the NIH staff for some share of the funds if the NIH has substantially more information on the respective productivity of the projects supported by that share of funds. Conversely, the legal restrictions imposed on any clinical research project, due to the involvement of human subjects, increase Congress' information about this type of research, and hence decrease Congress' incentive to delegate the allocation decision to the NIH. In Aghion and Tirole 1997's words, Congress has formal authority on the allocation of federal funds for basic research, and real authority on the allocation of federal funds for clinical research.

2 Related Literature

This study contributes to two strands of research: the empirical literature on the determinants of medical innovations and the literature on congressional bargaining. Several studies belonging to the former literature examine the impact of various sources of funding on the production of medical research (see e.g. Blume-Kohout, Kumar and Sood 2009, Jacob and Lefgren 2007 specifically for the NIH, and the whole field of scientometrics). Productivity of research varies substantially across sources of funding, which suggests to investigate the motivations of the institutions that invest in R&D. Following the profit incentives hypothesis, Azoulay and Tay 2003, Cerda 2003, Acemoglu and Linn 2004, Lichtenberg and Waldfogel 2003 and Finkelstein 2004, for instance, show that changes in the (potential) demand for a cure, or for a related good, induce changes in the nature or the volume of pharmaceutical innovations. Among them, Azoulay and Tay 2003, Lichtenberg and Waldfogel 2003 and Finkelstein 2004 use public health decisions or recommendations, regarding drugs for orphan diseases and vaccines, as exogenous shocks on the demand of medical goods to control for the productivity

of medical research. Such studies motivate the investigation of the factors driving public decisions themselves, which result from some bargaining among legislators who represent specific constituencies. Few papers study the factors that impact the amount of public funds devoted to research on some given topic. Lichtenberg 1999 finds a correlation between NIH funds and mortality, but does not control for scientific productivity. Bhattacharya and Packalen 2008 and Cutler, Meara and Richards 2009 link the level of investment in research to mortality in the US, the former using the same instruments as Acemoglu and Linn 2004. Here, I find that mortality in *certain* districts has a larger impact, and estimate this impact, controlling for research productivity. Like Ellison and Wolfram 2001 and Miller 2008, this paper stresses the role of political economy on medical issues. Under the assumption that congressmen affiliated to the groups studied here do not have distinct preferences for medical research, this paper also shows that the link between mortality and public funding of medical research results from a combination of congressional power and local concerns. In addition, this paper contributes to the open question of empirical effects of the composition of majorities in Congress (see e.g. Jayachandran 2006 and Snowberg, Wolfers and Zitzewitz 2006) and provides evidence that the House majority party has a substantial influence on the allocation of federal funds for research.

The local concerns studied here also differ from the local pork-barrel objectives usually raised in the political economy literature on congressional bargaining. Several studies (e.g. Cohen and Noll 1991, Ferejohn 1974, Atlas et al. 1995 and Knight 2005) show how the geographic origin of members of relevant committees affects the geographic allocation of federal funds. This influence has proven useful in explaining other differences across states (e.g. Aghion et al. 2005 or Levitt and Snyder 1997). For the NIH, Payne 1999, Payne 2003 and Hegde 2009 find some impact of congressional power on the geographic allocation of NIH funds. Indeed, Hegde 2009 finds that research centers located in the states represented in HouS do receive more research funds, but mitigates the impact of such earmarking by showing that this effect is limited to centers with relatively little expertise. The bulk of NIH funds traditionally goes to the same small set of universities, which do have expertise in any field, so that, on aggregate, earmarking to local research institutes should have limited consequences. Here, I find that the effect of local pork-barrel is indeed not significant at the state level.

3 The NIH Budget System

Institutionally, Congress and the President decide on the NIH budget every year. The Appropriations committees of both chambers are in charge of designing a budget proposal that will then be submitted to the whole Congress — the House of Representatives first, the Senate second.[5] The Appropriations bill for the NIH specifies the budget to be received by every institute (or center or agency) composing the NIH. After the budget is passed, these institutes, in turn composed of experts on medical research, are responsible for distributing this money through research grants to support research projects on topics related to their area of interest. The partition of the NIH itself results from congressional approval; a bill can create or terminate an institute or a center. For instance, Congress refused to create an entity specialized in the funding of research on HIV, proposed by President Clinton.[6] In other words, there is no institutional limitation on congressmen and the president's formal power over the distribution of public funds across research topics. In practice, congressmen thus have direct control over the research agenda of the NIH through two channels: deciding on the organization of the NIH and specifying the budget of every institute.

Some institutes are specifically responsible for supporting research on a single disease or set of diseases (such as the National Cancer Institute, or the National Institute of Allergy and Infectious Diseases), whereas others may award grants for basic research on any topic (such as the National Institute of General Medical Sciences) or support projects studying specific technology (such as the National Institute of Biomedical Imaging and Bioengineering). As a result, in several instances, the foci of the institutes and centers overlap, whereas some topics are strictly included in the foci of a single institute. In addition to controlling the Appropriations bill, congressmen may detail their preferences regarding the use of NIH funds. These views are recorded in the congressional reports. An examination of the reports suggests that both research productivity and distributive concerns determine the allocation of the NIH budget. The congressional report of the 104th Congress relative to the Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations bill contains the following statement:[7]

[5] Like the rest of the budget, it is enacted unless it is vetoed by the President.

[6] However, Congress approved the creation of the National Institute on Minority Health and Health Disparities (Source: AAAS R&D Funding Update December 18, 2000).

[7] 104th Congress, 2nd Session House of Representatives, 104-659, Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriation Report, 1997.

"The [Appropriations] Committee believes that NIH should distribute funding on the basis of scientific opportunity.[...] To enhance NIH's flexibility to allocate funding based on scientific opportunity, the Committee has attempted to minimize the amount of direction provided in the report accompanying the bill. [...] In stating that scientific opportunity should be the basis for allocating research funding, the Committee understands that other factors are also relevant to NIH's decisions, including such considerations as the infectious nature of a disease, the number of cases and deaths associated with a particular disease, the Federal and other costs of treating a disease, the years of productive life lost due to a particular disease, and the estimated proximity to research breakthroughs. The Committee does not presume to judge which criteria should take precedence in individual funding decisions, but urges NIH to consider the full array of relevant criteria as it constructs its research portfolio."

In addition, the bills often specify how the funds might be allocated across topics within the institutes or centers. As in the previous excerpt, the bills often repeat that the NIH has authority on the allocation of funds, yet give precise suggestions for the use of those funds. For instance, the previous report mentions:

"AIDS Funding. — Consistent with the philosophy outlined above, the Committee has again chosen not to earmark a specific dollar amount for AIDS research and has not provided a single appropriation for the Office of AIDS Research. In relying on NIH's recommendations for the allocation of the total funding provided by the Committee, the Committee understands that it would be NIH's intent to allocate AIDS funding in the following manner:[...]" [The report then details the amount of funds for research on HIV to be awarded by every institute or center.]

The Appropriations committees of either chamber (HouA and SenA), cited here, are not the only groups that may influence the appropriations of funds across research areas: within the Appropriations committees, the Subcommittees on Labor, Health and Human Services, Education and Related Agencies of the Appropriations Committee (HouS and SenS) have the specific task of drafting the NIH budget proposal; in the Senate, the Committee on Health, Education, Labor and Pensions (SenL) is in charge of providing recommendations for the NIH budget. Tables 1 and 2 summarize the list of groups I consider in this study, and present some information on their role and their size for every term of the period.

4 Description of the Data

I construct two sets of data for this study: data on mortality by districts or states represented in the congressional groups reported in table 1, and data on the NIH R01 grants. Both sets of data span the period 1985-2002, i.e. nine terms. Table 14, in the Appendix, reports summary statistics of all the variables used in this work.

4.1 Data on mortality and congressional groups

The first set of data merges three types of information on congressmen's affiliations from the congressional directories for the period 1985-2002, data from the death certificates that are recorded every year in the *Vital Statistics* data files, and data on the composition of congressional districts, provided by Adler 2008.[8] For every death, the *Vital Statistics* data files report the age, the main cause of death, the state of residence of the deceased, and whether the deceased was Black or not. I use these data to compute the number of life-years lost to 100 years old, $life-years_d^s$, in the state represented by the senator s, because of the disease d in a given year:[9]

$$life-years_d^s \equiv \sum_{\substack{i,\ resident\ of \\ the\ state\ of\ s}} (100 - Age\ of\ death_i)\mathbf{1}_{\{i\ died\ of\ d\}} \tag{1}$$

Since the *Vital Statistics* files do not specify the district of the deceased, I use information on the race of congressional districts residents from Adler 2008 to estimate the number of life years lost $life-years_d^r$ (plus 100) in the district represented by the representative r, because of the disease d:

$$\begin{aligned} life-years_d^r \equiv\ & \pi_{bl}(r,s) \sum_{\substack{i,\ resident\ of \\ the\ state\ of\ r}} (100 - Age\ of\ death_i)\mathbf{1}_{\{i\ died\ of\ d\}} \times \mathbf{1}_{\{i\ Black\}} \\ & + \pi_{nbl}(r,s) \sum_{\substack{i,\ resident\ of \\ the\ state\ of\ r}} (100 - Age\ of\ death_i)\mathbf{1}_{\{i\ died\ of\ d\}} \times \mathbf{1}_{\{i\ non-Black\}} \end{aligned} \tag{2}$$

where $\pi_{bl}(r,s)$ (resp. $\pi_{nbl}(r,s)$) is the share of Black (resp. non-Black) residents of the state of senator

[8] These latter data are available on http://sobek.colorado.edu/~esadler/Congressional_District_Data.html.

[9] Using the difference between the age at death and some other high age, e.g. 120 years old, gives very similar results in the following estimations.

s living in the district represented by the representative r.[10]

With the previous notations, I define $LY^g_{d,t}$, the share of life-year lost due to a disease d in all the districts (resp. states) represented in the group g of the House (resp. Senate) in Congress term t, as:[11]

$$LY^g_{d,t} \equiv \frac{\sum_{c \in g \ at \ t} life - years^c_d}{\sum_{d \in D} \sum_{c \in g \ at \ t} life - years^c_d} \tag{3}$$

where D is the set of diseases. In some specifications, I also use the *average* number of life-years lost in the districts of the group g:

$$\ell y^g_{d,t} \equiv \frac{\sum_{c \in g \ at \ t} life - years^c_d}{\#\{c \in g \ at \ t\}} \tag{4}$$

I use the *Vital Statistics* file for the year 1991 to estimate life-years lost variables for all the groups of Table 1 and all the Congress terms. The relative impact of a disease on a state is extremely stable over the period, so that using mortality data of 1984 or 1998 gives the same results in the estimations.[12]

Table 3 reports summary statistics of the share of life-years lost by disease for the US as a whole in 1991, and for the House majority (HouM). The set of twelve diseases (or disease categories) retained in this study causes more than 90 per cent of the deaths in the US in 1991. All the other deaths are due to accidents, murders, or unidentifiable causes. The table shows the correlation between the distribution of life-years lost across diseases in the US and in HouM or HouS. The standard deviation of life-years lost in HouM and HouS reflect changes in the composition of these groups every Congress term. Figure 1 illustrates the change of mortality on the House majority following the 1995 majority shift.

4.2 Data on NIH grants

The second set of data derives from detailed information on R01 grants at the grant level provided in various NIH databases: the *REsearch Portfolio Online Reporting Tools (REPORT)* for the period 1992-2009, the NIH data supplied by the NBER (Lichtenberg 1999) for the period 1986-1991 and the

[10] Since Adler 2008's data do not contain information on the last years of the period, I assume that the composition of the districts did not changed between 1998 and 2002 to compute life-years lost in districts after 1998.

[11] I count twice the life years lost of a state whose two senators are in the group.

[12] These estimations are available upon request.

Computer Retrieval of Information on Scientific Projects (CRISP) database.[13] I use these data to gather the following information for every competing R01 grant: the fiscal year the grant was awarded (which is by definition one year after the budget is voted on), the total amount of funds received by the grantee for the whole length of the grant, the state where the research institution of the grantee is located, the specific agency within the NIH awarding the grant and a set of keywords describing the research project funded by the grant.

The keywords describing a grant mention whether the research hereby funded involved human subjects. I define such research as *clinical*, and the other projects as *basic*.

When possible, I also match every grant to a unique disease in the set of disease or disease categories retained here (see Table 3). The matching is not possible for some grants that support research on sets of conditions that are too large to be matched with a unique disease (e.g. *pathology*), or that are related to non-lethal conditions (e.g. *back injury*).[14] Finally, due to missing information, some grants cannot be matched to a US state. Table 13, in Appendix, reports the sum of funds awarded through R01 grants by Congress term (column "Total funds"). After removing all the missing observations and the grants that cannot be matched to a single disease, I obtain a sample that amounts to more than 80 per cent of the actual budget for R01 grants on average (column "Sample funds").

For the purpose of this paper, I aggregated these data so as to obtain the total amount of funds awarded by disease, by type of research and by Congress term (and by state, for some specifications). Table 3 reports some summary statistics by disease and type of research. They show the positive correlation between the funds for research on a disease, either basic or clinical, and the share of life-years lost, in the US or in the House majority.

[13] Except for the NBER data, all the other data are now available on http://report.nih.gov/. The NIH documentation, available on http://grants.nih.gov/grants/funding/r01.htm, mentions: "The Research Project Grant (R01) is the original and historically oldest grant mechanism used by NIH. The R01 provides support for health-related research and development based on the mission of the NIH. [...] The Research Project (*R*01) grant is an award made to support a discrete, specified, circumscribed project to be performed by the named investigator(s) in an area representing the investigator's specific interest and competencies, based on the mission of the NIH."

[14] The matching strategy, partly based on Toole 2007, and the other details of the construction of these data are described in a thorough technical appendix available upon request.

5 Model and Empirical Strategy

5.1 Congress' Allocation Problem

This section presents a simple static model that aims to formalize the description of the NIH budget process, and provides a basis for the discussion in Section 7.[15] Let K denote the set of "research categories" partitioning the NIH budget. For instance, K could be the set of diseases, the set of research institutes, the combination of both, or any other set relevant for the classification of research projects. There are two agents: Congress C and the NIH staff N. For any research category $k \in K$, Congress C may choose a level of public investment $F_k \in \mathbb{R}^+$. Investing F_k in research k yields a scientific return equal to $c_k F_k^\gamma$, — for example the "production of knowledge on topic k" — where $\gamma \in (0,1)$ and the productivity factor c_k is independently drawn from a random variable with density $g_k(c_k)$, with a support in $\mathbb{R}^+$.

For investments $(F_k)_{k\in K}$, the ex-post utility derived by $i = C$ or N is:

$$U^i\Big((F_k)_{k\in K}\Big) \equiv \sum_{k\in K} c_k F_k^\gamma A_k^i - \lambda^i \sum_{k\in K} F_k \tag{5}$$

where $\forall k \in K, A_k^i > 0, \lambda^C > 0, \lambda^N = 0$. I assume that, ex-ante, Congress C does not know c_k, whereas N does, but that C has the possibility to delegate the allocation of a chosen amount of funds into the categories of a chosen subset M of K to the NIH N, and decides of the amount of funds for the other categories $L \equiv K \setminus M$. Congress C's problem is then:

$$\begin{cases} Max_{\{F\in\mathbb{R}^+, L\subset K, \mathbf{F}_L\in\mathbb{R}^{+\ \#L}\}} \mathbf{E}\left[\sum_{k\in L} c_k F_k^\gamma A_k^C + \sum_{k\in M} c_k \hat{F}_k^\gamma A_k^C \middle| \mathbf{g}\right] - \lambda^C F \\ s.t. \begin{cases} \hat{\mathbf{F}}_M \equiv Argmax_{\{\mathbf{F}_M\in\mathbb{R}+^{\#M}\}} \sum_{k\in M} c_k F_k^\gamma A_k^N \\ \sum_{k\in M} F_k = F - \sum_{k\in L} F_k \end{cases} \\ \sum_{k\in L} F_k + \sum_{k\in M} \hat{F}_k = F \\ L \cup M = K, L \cap M = \emptyset \end{cases} \tag{6}$$

where: $\mathbf{F}_L \equiv (F_k)_{k\in L}$ for any $L \subset K$ and $\mathbf{g} \equiv (g_k)_{k\in K}$. A solution to that problem always exists, since a solution of the sub-problem defined by adding the constraint $L = L_0$ to the problem 6 has

[15] Although it also substantiates some specifications, it is not necessary for the understanding of most of the empirical sections 4, 5.2 and 6.

a solution for any $L_0 \subseteq K$, and there is a finite number of such subsets. The solution need not be unique, however. Let $\left(F^*, L^*, \mathbf{F}^*_{L^*}\right) \in \mathbb{R}^+ \times 2^K \times \mathbb{R}^{+\ \#L^*}$ denote a solution of the previous program.

Proposition 1. *Let $K \equiv \{1, ..., n\}$ the set of research categories, $g_k(c)$, the probability density function of the productivity factor c_k, and assume that g_2 second-order dominates g_1 and $A_1^i = A_2^i$, for $i = C, N$. If $1 \in L^*$ for some solution $\left(F^*, L^*, \boldsymbol{F}^*_{L^*}\right) \in \mathbb{R}^+ \times 2^K \times \mathbb{R}^{+\ \#L^*}$ of the Congress' problem, then there also exists a solution $\left(\tilde{F}^*, \tilde{L}^*, \boldsymbol{F}^*_{\tilde{L}^*}\right) \in \mathbb{R}^+ \times 2^K \times \mathbb{R}^{+\ \#\tilde{L}^*}$ of this problem such that $2 \in \tilde{L}^*$.*

Proof. See Appendix. □

The proposition means that C is more likely to delegate the allocation of funds across research categories about which it has less precise information.

Specification of the functional forms — For the largest level of aggregation in the estimations, the set K is $D \times \Theta$, where D is the set of diseases and $\Theta \equiv \{basic, clinical\}$ the set of types of research. The main variable of interest at this level of aggregation is the log of the share of life-years lost due to disease d in the group g, denoted $\log LY_d^g$. Given the set of potentially powerful groups G in Congress, I posit the following reduced form for A_k^C (expressed in logs for clarity):

$$\log A_{d,\theta}^C = \sum_{g \in G} w^g \times (\pi^g + \pi^g_{Clinic} \times Clinic_\theta) \log LY_d^g \tag{7}$$

where w^g is the institutional bargaining power of group g, π^g and π^g_{Clinic} are two parameters that allow the utility derived from medical research to vary across types of research, and $Clinic_\theta$ is a variable equal to 1 if $\theta = clinical$ and 0 otherwise.[16] Since part of the allocation decision may be delegated to the NIH staff, the first-order conditions of C and N's problems imply that the amount of funds $F_{d,\theta}$, awarded for research of type θ on disease d, can be written:

$$\begin{aligned} \log F_{d,\theta} \quad = \quad & \sum_{g \in G} \frac{\delta_{d,\theta} \times w^g \times (\pi^g + \pi^g_{Clinic} \times Clinic_\theta)}{1-\gamma} \times \log LY_d^g + \\ & \frac{1-\delta_{d,\theta,s}}{1-\gamma} \log A^N_{d,\theta} + \frac{\log \gamma - \delta_{d,\theta} \log \lambda^C}{1-\gamma} + \frac{\delta_{d,\theta} \log \mathbf{E}[c_{d,\theta}] + (1-\delta_{d,\theta}) c_{d,\theta}}{1-\gamma} \end{aligned} \tag{8}$$

[16] Note that congressmen need not be partitioned into groups, but may belong to several of them.

where $\delta_{d,\theta}$ is equal to 1 if the amount of funds allocated to the "research category" (d,θ) is decided by Congress, i.e. $(d,\theta) \in L*$ with the previous notations, and 0 if it is decided by the NIH. Consider the products of parameters of the first term of the right-hand-side in the equation 8, renamed β^g and β^g_{Clinic}:

$$\begin{aligned} \beta^g &\equiv \frac{1}{1-\gamma} \times \delta_{d,basic} \times w^g \times \pi^g \\ \beta^g_{Clinic} &\equiv \frac{1}{1-\gamma} \times \delta_{d,clinical} \times w^g \times \pi^g_{Clinic} \end{aligned} \tag{9}$$

For the finest level of aggregation in the estimations, the set K is $D \times \Theta \times S$, where S is the set of American states. The variables of interest at this level of aggregation are the number of congressmen of the state s in the group g, denoted NC^g_s, and $\log LY^g_d$. With similar notations as before, the amount of funds $F_{d,\theta}$ for research of type θ on disease d awarded to research institutes located in the state s can be written:

$$\begin{aligned} \log F_{d,\theta,s} = & \sum_{g \in G} \frac{\delta_{d,\theta,s} \times w^g \times (\pi^g_1 + \pi^g_{Clinic,1} \times Clinic_\theta)}{1-\gamma} \times \log LY^g_d + \\ & \sum_{g \in G} \frac{\delta_{d,\theta,s} \times w^g \times (\pi^g_2 + \pi^g_{Clinic,2} \times Clinic_\theta)}{1-\gamma} \times NC^g_s + \\ & \frac{1-\delta_{d,\theta,s}}{1-\gamma} \log A^N_{d,\theta,s} + \frac{\log\gamma - \delta_{d,\theta,s} \log \lambda^C}{1-\gamma} + \frac{\delta_{d,\theta,s} \log \mathbf{E}[c_{d,\theta,s}] + (1-\delta_{d,\theta,s}) c_{d,\theta,s}}{1-\gamma} \end{aligned} \tag{10}$$

I rename the products of interest in this equation β^g, β^g_{Clinic}, ρ^g and ρ^g_{Clinic} as follows:

$$\begin{aligned} \beta^g &\equiv \frac{1}{1-\gamma} \times \delta_{d,basic} \times w^g \times \pi^g_1 \\ \beta^g_{Clinic} &\equiv \frac{1}{1-\gamma} \times \delta_{d,clinical} \times w^g \times \pi^g_{Clinic,1} \\ \rho^g &\equiv \frac{1}{1-\gamma} \times \delta_{s,basic} \times w^g \times \pi^g_2 \\ \rho^g_{Clinic} &\equiv \frac{1}{1-\gamma} \times \delta_{s,clinical} \times w^g \times \pi^g_{Clinic,2} \end{aligned} \tag{11}$$

In the estimations, I assume that the parameter w^g reflects the institutional rules that govern congressional decisions, which have not changed over the period studied. Equations 8 and 10 show that I can infer the bargaining power w^g of any group g, up to a constant normalization factor, from the estimation of β^g, β^g_{Clinic}, ρ^g_{Clinic} or ρ^g_{Clinic}, under the assumption that the parameter π^g is constant across groups, i.e. no group has a distinct interest for medical research. I discuss more this point in Section 5.2.

5.2 Empirical strategy and specification

Consider a group g of interest, $LY_{d,t}$ the share of life-years lost due to disease d in the districts represented in the group g at t, formally defined in equation 3, and $NC_{s,t}$ the number of congressmen of state s in g at t (the superscript $.^g$ on these variables, present in the model of Section 5.1, is omitted here for readability.) The main purpose of the empirical part is to estimate the effect of these two variables on the allocation of public funds for medical research (i.e. the parameters β and β_{Clinic} and/or ρ and ρ_{Clinic} in equations 9 and 11 in the Section 5.1).

To do so, at the largest level of aggregation, I estimate a linear model of the form:

$$\log F_{d,\theta,t} = c_{d,\theta} + \beta \log LY_{d,t} + \beta_{Clinic} \log LY_{d,t} \times Clinic_\theta + \mu X_{d,\theta,t} + \epsilon_{d,\theta,t} \tag{12}$$

And, at the finest level of aggregation, I estimate a linear model of the form:

$$\log F_{d,\theta,s,t} = c_{d,\theta,s} + \beta \log LY_{d,t} + \beta_{Clinic} \log LY_{d,t} \times Clinic_\theta + \rho NC_{s,t} + \rho_{Clinic} NC_{s,t} \times Clinic_\theta + \mu X_{d,\theta,s,t} + \epsilon_{d,\theta,s,t} \tag{13}$$

where $F_{d,\theta,t}$ (resp. $F_{d,\theta,s,t}$) is the amount of funds for research on disease d of type θ awarded in Congress term t (resp. on disease d of type θ, awarded to research institutes in state s in Congress term t), and $Clinic_\theta$ is a dummy variable equal to one if $\theta = clinical$, and 0 $\theta = basic$. The vector of variables $X_{d,\theta,t}$ (resp. $X_{d,\theta,s,t}$) may include control variables such as LY or NC variables for other groups, as well as Congress term dummy variables, and their interactions with $Clinic_\theta$. The term $c_{d,\theta}$ (resp. $c_{d,\theta,s}$) is the unobserved fixed effect for research on disease d of type θ (resp. on disease d of type θ performed in state s). These terms correspond to all the terms on the lowest lines of equations

8 and 10. Standard errors are clustered at the disease and type of research level (resp. at the disease and type of research and state level).

For the estimated coefficients of interest $\hat{\beta}$, $\hat{\beta}_{Clinic}$, $\hat{\rho}$ or $\hat{\rho}_{Clinic}$ to represent the group g's bargaining power as in equations 9 and 11 of Section 5.1, the two following necessary assumptions must be satisfied: (1) the productivity of research on a disease is not correlated with the relative impact that disease has on the constituents represented in the group g, and (2) the utility congressmen derive from medical research is uncorrelated with their affiliation to that group.

The exogenous timing of congressional elections supports the identification assumption (1): every two years, the composition of the majority (or of any committee or subcommittee) may change, following elections outcomes. The affiliation of a congressman to the majority party is a direct consequence of the elections, and the assignment of a congressman to a committee or subcommittee is decided within her/his party at the beginning of every term. A congressman's assignment will depend on her/his preferences and clout at the time of the process, as well as on congressional rules governing the assignment process (Schneider 2007). The life-years lost function of a group induced by the elections is thus likely uncorrelated with the distribution of shocks on research productivity across diseases.

The identification assumption (2) requires that congressmen do not enter the group only because they have a distinct interest in medical research, or because they felt wronged in the previous NIH budget decisions. Assumption (2) should hold for members of any party, or of any Appropriations committee, since they can be involved in any decision regarding the use of federal funds, not only the budget for medical research. This assumption is more problematic for the subcommittees HouS and SenS and for the committee SenL, however, even though these subcommittees and this committee are also in charge of several issues such as education or labor conditions.

To account for the possible failure of assumption (2), I conjecture that, within the Appropriations committees, congressmen with more political influence are assigned to the subcommittees with a higher average share of budget by member. Two facts support this conjecture: the size of the subcommittees is constrained by congressional and party rules (Schneider 2007) that are stable throughout years, and the total size of the budget that falls upon the responsibility of any subcommittee seems roughly exogenous (see for instance Dalton 2000's report of the doubling of the NIH budget starting in 1996). I use the number of consecutive terms a congressman c has served in the same chamber ($tenuC_{c,t}$)

and in the Appropriations committee ($tenuA_{c,t}$), counting backwards from t, as proxies for political influence. For any congressman $c \in C$ active at t, and any subcommittee g, I estimate the probability for congressman c to be affiliated with the group g, given that she or he has served $tenuC_{c,t}$ successive terms in the chamber, and $tenuA_{c,t}$ successive terms in its Appropriations committee:

$$P(c \in g, t|tenuA_{c,t}, tenuC_{c,t}) = \Phi(\omega_0 \frac{1}{tenuC_{c,t}} + \omega_1 tenuA_{c,t} + \mu X_t) \tag{14}$$

where X_t is a vector of Congress term dummies. Denoting the predicted probabilities

$$\hat{P}(c \in g, t|tenuA_{c,t}, tenuC_{c,t}),$$

the instrument for the first difference of the log of $LY^g_{d,t}$ is the first difference of the log of:

$$LY^{g*}_{d,t} =\equiv \frac{\sum_{c \in g \ at \ t} \hat{P}(c \in g, t|tenuA_{c,t}, tenuC_{c,t}) \times life - years^c_d}{\sum_{d \in D} \sum_{c \in g \ at \ t} \hat{P}(c \in g, t|tenuA_{c,t}, tenuC_{c,t}) \times life - years^c_d} \tag{15}$$

The same procedure can also be used to construct instruments for the number of congressmen of a state in HouS or SenS, or instruments for life-years lost in the states represented in SenS or SenL.

6 Results

6.1 Allocation of funds across diseases

This section focuses on the impact of life-years lost across diseases and groups on the allocation of funds across diseases, i.e. on the estimation of the coefficients β and β_{Clinic} of equations 12 and 13 for the set of groups of table 1.

Table 4 reports the OLS estimation of the coefficients of the first difference of equation 12, separately for every group, twice.[17] The first series of estimations (columns 1 to 7) shows that life-years lost in HouS impact NIH funds for clinical research, positively and significantly. No other group, except HouM (significant at 10%), has a significant positive impact on the budget. The second series of estimations (columns 8 to 13), indicate that the effect of HouS holds when controlling for other groups. In all the estimations, life-years lost in HouS have a negative impact on basic research. This

[17] The small size of the sample and the correlation between life-years lost variables would limit the interpretation of a regression including all life-years lost variables at once.

effect is significant at 10%, for some specifications. The estimated negative effect of life-years lost in HouS on NIH funds may reflect HouS' smaller bargaining power on the allocation of basic funds. It may also result from the negative correlation of life-years lost in HouS with some omitted variables. The previous tables show that the significance of that term disappears when life-years lost variables for other groups are included as well, which support that latter hypothesis. Overall, no group has any robust impact on the allocation of funds for basic research.

Depending on whether we take into account the estimated coefficient $\hat{\beta}$, a 1 percent increase in life-years lost in districts represented in HouS leads to a 1.3-2.7 percent increase in funds for clinical research. Figure 2 illustrates the previous effects. It represents the net increase of funds (in %) for basic and clinical research respectively, with regard to the net increase of life-years lost in the districts represented in HouS, separately for every Congress term, and the linear fit of the resulting set of points. There are almost as many graphs showing a positive (4 graphs) and a negative (3 graphs) slope for the linear fit of the relationship between life-years lost and funds for basic research, which suggests that (at least the sign of) the correlation between basic research funds and life-years lost is null. However, for any Congress term, except possibly for the term starting in 1997, the slope of the linear fit of the plots representing the increase of funds for clinical research is greater than the linear fit of the plots representing the increase of funds for clinical research with respect to increases in life-years lost. In addition, these figures show that the estimations do not stem from the behavior of some obvious disease outlier.

The previous specifications are more suited to the actual formulation of the budget described in section 3. However, as a robustness check, I estimate the parameters of equation 13 in Table 5, with robust standard errors clustered by disease, type of research and state.[18] These estimations confirm the effect of life-years lost in districts of HouS members on clinical research and the absence of effect of this group on the funds for basic research. This effect is robust to the inclusion of variables $NC_{s,t}$, which control for the potential earmarking of research funds to the representatives' states. (The estimated coefficients of $NC^g_{s,t}$ are reported later for clarity.) In addition, at this level of aggregation, life-years lost in HouA and HouM have a positive and significant impact on research funds for clinical research for some specifications. By construction, if states are similar and congressmen do not directly

[18] As mentioned in the Section 4, I assume that the missing observations (less than 200 out of 9600) result from missing information on the amount of funds for research on a given disease of a given type in a given state. A Tobit estimation would give similar results.

earmark NIH funds to some states, this level of aggregation leads to lower estimated standard errors than when data are aggregated by disease and type of research only. States, however, are not similar, and the skewness of the distribution of NIH funds across states could explain part of the new results. To check this possibility, I estimate the coefficients of a variant of equations 12 and 13, replacing the log of funds by the funds themselves, and the log of life-years lost by the average number of life-years lost in the districts represented in the group of interest (ℓy in equation 4). Table 6 reports the results of these estimations. They suggest that the composition of HouM does have a substantial impact on the amount of funds by disease and by type of research and by state, whereas the estimated coefficients for HouA are rather small — none of these coefficients is significant, though. The large estimated coefficients of "life-years lost in HouA" variables thus reflect changes of funding for research of a certain type, on certain diseases and in certain states that represent a negligible part of the NIH budget. These estimations also support the role of HouS on the funds allocation. Given that HouS contains around 14 members, columns 4 or 6 in Table 6 show that an additional life-year lost due to a disease in HouS leads to an increase of NIH funds by year for clinical research on that disease equal to around $\frac{102{,}000}{2\times14} \simeq \$3600$ by year.

Table 7 displays the results of the 2SLS estimation of equations 12 and 13, using the instrument defined in equation 15. Life-years lost in HouS affect the allocation of NIH funds across diseases, as long as life-years lost variables of HouM are not included in the controls. The correlation between HouM and HouS districts' characteristics, stemming from the fact that the share of majority members in HouS is equal to the share of majority members in the whole House, limits the possibility of distinguishing the respective effects of life-years lost in HouS and HouM in a 2SLS estimation (note that, even though these estimations are not shown here, controlling for life-years lost variables of any group other than HouM does not affect the significance of the effect of life-years lost in HouS' districts). These results may thus either confirm that HouS members have a distinct bargaining power in the congressional negotiations over the NIH budget, or suggest that members of the House majority also have some bargaining power on the allocation process. I explore this latter hypothesis in Section 6.2.

6.2 The role of the House majority

The estimated influence of HouS and HouM suggest that congressmen who belong to both group may have a large bargaining power. To estimate this power, I split the HouS into two groups, congressmen who belong to HouM, and those who do not. The estimations, reported in the first columns of Table 8, show that members of the minority of HouS indeed have no impact on the allocations of funds across diseases. In fact, since committees' and subcommittees' affiliations are decided within parties, the influence of a district on the budget stems from the influence of its representatives *in a party* and the bargaining power of its representative in Congress. Testing this hypothesis would require knowing the weight that every district, through its representatives in a party, has on that party's objectives. Since I do not observe these weights, I define an ad hoc "party's life-years lost function" for the Democratic and Republican parties that is the weighted sum of life-years lost in US districts, where the weight of a district is the number of times that district's House representative belonged to that party. The share of life-years lost in the majority party is then, for the terms starting between 1985 and 1993,

$$LY_{d,t} \equiv \frac{\sum_{y=1985}^{1993} \sum_{\substack{c \ in \ g, \\ at \ y}} life - years_d^c}{\sum_{d \in D} \sum_{y=1985}^{1993} \sum_{\substack{c \ in \ g, \\ at \ y}} life - years_d^c} \tag{16}$$

and, for the terms starting between 1995 and 2001,

$$LY_{d,t} \equiv \frac{\sum_{y=1995}^{2001} \sum_{\substack{c \ in \ g, \\ at \ y}} life - years_d^c}{\sum_{d \in D} \sum_{t=1995}^{2001} \sum_{\substack{c \ in \ g, \\ at \ y}} life - years_d^c} \tag{17}$$

The majority party for any term starting between 1985 and 1993 (resp. between 1995 and 2001) is thus the fictitious group of representatives active between 1985 and 1994 (resp. between 1995 and 2002), in which a representative appears n times if she was n times in HouM between those years. This definition assumes that the more "loyal" districts have more influence in the party. Columns (3), (5) and (7) of Table 8 report the estimation of the effect of the share of that life-years lost function on the NIH funds. They show that life-years lost in the districts that have been more loyal to the majority party of the House have a significant and positive effect on the allocation of funds for clinical research across diseases. They do not have any impact on the allocation of funds for basic research,

except in Column (4), which only suggests that HouM and the fictious "majority party" are very similar. To summarize these last results and the previous ones, some districts have a larger weight in the Congressional allocation process, this larger weight stems from their representatives' membership in the majority party, or from their House representatives *possible* membership in HouS, or both — the interpretation of the estimation results depending on the validity of the identification assumptions.

To illustrate the consequences of the weight of the House majority party, I estimate the counterfactual allocation of funds across diseases that would have occurred if the Republican party had not won the majority between 1995 and 2002. The results of the counterfactual analysis are reported in Table 9. They show a substantial increase in public funding for clinical research on psychiatric disorders, central nervous system diseases and respiratory disorders due to the majority change, and a substantial decrease of funding for clinical research on congenital, hypertensive and HIV-related disorders.

6.3 Allocation of funds across states

Table 10 reports the estimations of the coefficients ρ of the equation 13 and some variants. The number of representatives of a state in HouS shows no significant positive effect on the research funds for that state, regardless of whether other groups' variables are included in the regressions. (In fact, it seems to have a negative impact on the funds by state; this result, however, is not robust to the inclusion of additional controls). No group seems to have any significant positive effect on the distribution of funds across states. These results do not imply that earmarking does not exist at all — in fact some empirical studies and anecdotal evidence suggest the opposite (see Section 2) — but they suggest that earmarking either affects the redistribution within states or involves relatively small amounts of funds.

The lack of any significant effect is quite surprising nevertheless, especially for the specifications that do not include the social cost variables (e.g. Column 1 of table 10), since one may conjecture that the research interests of research institutions are correlated with the health conditions of their neighboring population. To investigate further that conjecture, I compute the log of the share of life-years lost $LY^s_{d,t} \equiv \frac{life-years^s_d}{\sum_d life-years^s_d}$ separately for every state, and regress the log of funds on that variable, as in:

$$\log F_{d,\theta,s,t} = \omega \log LY^s_{d,t} + \omega_{Clinic} \log LY^s_{d,t} \times Clinic_\theta + X_{d,\theta,s,t} + u_{d,\theta,s,t} \tag{18}$$

where $X_{d,\theta,s,t}$ may include dummy variables for every Congress term, type of research, disease and state, and the interactions between the type of research and other dummy variables. The results (see Table 11) show no significant impact of life-years lost on research funds at the state level. Figure 3 represents the funds for research in a state with respect to life-years lost in that state for every disease, aggregated over all Congress terms and over both types of research. It confirms the result of the last estimations and shows that there is no systematic positive correlation between research interests and local health conditions. In fact, there is no obvious positive correlation between research interests and local health conditions for any disease other than HIV, infectious disorders and hypertension. Once aggregated at the state level, committee membership thus has no significant effect on the geographic allocation of NIH funds, either directly or indirectly through the allocation of funds by disease.

7 Discussion

The previous results raise at least two immediate questions: what causes a group's bargaining power, and why does the composition of the majority or HouS have more effect on clinical than basic research?

What causes a group's bargaining power? As in the seminal paper of Baron and Ferejohn 1989, being the proposer of the NIH budget to Congress can provide some additional bargaining power, if congressmen are impatient. This mechanism may explain the power of HouS, although the Senate committee SenL also makes recommendations on the NIH budget. In fact, on a different question, Knight 2005 finds that it is the Transportation Authorization Committee, roughly the equivalent of SenL for transportation issues, that has power on the distribution of funds for road projects. In addition, the estimations show no effect of the composition of SenS, the equivalent of HouS in the Senate. The question of the specific influence of HouS among all potential proposers thus remains open. This study also leaves open the question of congressmen's motivations. Are they interested in their constituents' health conditions for electoral purposes, or are they concerned with their own health conditions or their relatives', which happen to be aligned with their constituents?

Differences between basic and clinical research. The results of this study do not indicate precisely whose diseases enter a congressman's utility. However, the fact that the composition of the House subcommittee affects only the allocation of clinical research, with no significant impact on basic research, suggests that congressmen are concerned with the productivity of research across diseases. To support this claim, I examine three differences between basic and clinical research that could explain this result, the third one being the most plausible cause of the estimation results. First, clinical research on a disease may be more likely to be located in states or districts relatively more affected by that disease. Congressmen would then use clinical research topics as an earmarking device towards research institutes located in their districts or states. But the estimated coefficients on the number of congressmen of a state in HouS or HouM, reported in column 2 of table 10, are not significant, which suggests the rejection of that first hypothesis. Second, the returns from investment in basic research (or the parameter π^g in equation 7 of Section 5.1) may be very small for congressmen: clinical research may be easier to publicize for electorally-motivated politicians, or clinical research could lead to faster pharmaceutical innovations. Small returns on basic research would not prevent congressmen from ranking research projects according to the social benefits they may bring (unless the returns from basic research were null in which case the funds for basic research should be null as well.) We should thus estimate $\hat{\beta} > 0$ in the estimations. However, this estimated coefficient is never large or significant. Third, the productivity of clinical research may be more easily estimated by non-specialists of medical research than basic research. This difference could stem from the complexity of scientific knowledge as well as from the legal restrictions that bind the uncertainty of clinical research, which, by definition, uses human data. If congressmen' s information on basic research is poor enough, Proposition 1 claims that they should delegate the allocation of these funds to the NIH staff, who shares congressmen's concern for research productivity, but have no reason to take into account the specific health conditions of powerful congressmen' s electorate.[19] For that claim to hold, the shocks on the productivity of research (such as the factors c_k in Section 5.1) should necessarily have a larger variance for basic than for clinical research. To investigate that point, I regress the squared residuals of the last regression shown in Table 4 on the dummy variable for clinical research $Clinic_\theta$. The estimated coefficient of $Clinic_\theta$, reported in Table 12, is negative, and significant when I include a dummy variable for every disease and/or Congress term. Under the assumption that the residuals are indeed shocks on

[19] NIH officials are usually appointed for several years.

research productivity, these regressions thus support the third hypothesis. Anecdotal evidence from the description of the NIH organization also supports that hypothesis: the explicit mission of one agency, the *National Institute of General Medical Sciences*, is to fund basic research on any disease. The allocation of the budget of this agency across diseases is thus effectively delegated to its staff.[20] As argued by Foucault 1963, political authority influences the focus of medical research, at least through the funding of this research. However, the concern for scientific productivity limits the effect of this authority on basic research.

8 Summary and Extensions

This paper finds that the social cost of a disease, measured in life-years lost, on the constituents represented either in the majority party of the House or in the subcommittee of the House Appropriations Committee in charge of health issues has a positive impact on the NIH funds for clinical research on that disease awarded through R01 grants. In addition, there is no significant effect of the state origin of the members of these groups on the allocation of funds across states.

These results show that congressional elections are a source of exogenous changes of NIH funding priorities. This source could be used to understand better the complementarity between public and private research (David, Hall and Toole 2000, Hall, Mairesse and Mohnen 2010, Toole 2007), and their joint effect on medical innovations. These results also contribute to the existing literature on Congress by raising another channel of committee power than the usual pork-barrel. The issue at stake here is not the pure distribution of private goods among congressmen, already considered in the legislative bargaining literature. It is not an ideological issue either, and, as such, requires another approach. Given that a large part of the public budget is distributed through agencies, the economic effect of these institutions demands closer theoretical and empirical investigation.

[20] Congressional bills may yet give some directions for the use of that budget. Moreover, a substantial number of basic grants are awarded by other agencies or smaller subdivisions of the NIH.

References

Acemoglu, Daron (2002), "Directed Technical Change," *Review of Economic Studies*, LXIX, 781-810.

Acemoglu, Daron and Joshua Linn (2004), "Market Size In Innovation: Theory And Evidence From The Pharmaceutical Industry", *Quarterly Journal of Economics,* Vol. 119, No. 3, pp. 1049-1090.

Adler, E. Scott (2008), "Congressional District Data File, 1991-1997." University of Colorado, Boulder, CO.

Aghion, Philippe, Leah Boustan, Caroline Hoxby and Jerome Vandenbussche (2005), "Exploiting States' Mistakes to Identify the Causal Impact of Higher Education on Growth", NBER Working Paper. Forthcoming, 2005.

Aghion, Philippe, and Peter Howitt (1992), "A Model of Growth through Creative Destruction," *Econometrica*, LX, 323-351.

Aghion, Philippe and Jean Tirole (1997), "Formal and Real Authority in Organizations", *Journal of Political Economy*, Vol 105 No 1.

Arrow, Kenneth (1962), "Economic Welfare and the Allocation of Resources for Invention." in Nelson, R. (Ed.), *The Rate and Direction of Inventive Activity.* Princeton, NJ.

Atlas, Cary, Thomas Gilligan, Robert Hendershott, and Mark Zupan "Slicing the federal government net spending pie: who wins, who loses , and why", *American Economic Review,* 85.

Azoulay, Pierre and Abigail Tay (2003), "Medical Progress and Health Care Financing: Research in Academic Medical Centers Following the 1997 Medicare Cuts", mimeo.

Baron, David and John Ferejohn (1989), "Bargaining in Legislatures", *American Political Science Review*, LXXXIII, 1181-1206.

Bhattacharya, Jayanta and Mikko Packalen (2008), "Is Medicine an Ivory Tower? Induced Innovation, Technological Opportunity, and For-Profit vs. Non-Profit Innovation", NBER working paper #13862.

Blume-Kohout, Margaret E., Krishna B. Kumar and Neeraj Sood (2009), "Federal Life Sciences Funding and University R&D", NBER working paper #15146.

Cerda, Rodrigo (2003), "Drugs, Market Size and Population" University of Chicago, Ph.D. thesis.

Cohen, Linda and Roger Noll (1991), *The Technology Pork Barrel*, Brookings Institution Press.

Cutler, David, Ellen Meara and Seth Richards (2009), "Induced Innovation and Social inequality: Evidence from Infant Medical Care", NBER working paper #15316.

Dalton, Rex (2000), "Clinton proposes $2.8 billion increase in science funding" *Nature* 403, 349 (27 January 2000)

David, Paul A., Bronwyn H. Hall and Andrew A. Toole (2000), "Is public R&D a complement or substitute for private R&D? A review of the econometric evidence", *Research Policy* 29, 497-529.

Drandakis, Emmanuel and Edmund Phelps (1966), "A Model of Induced Invention, Growth and Distribution", *The Economic Journal*, December.

Fisher Ellison, Sara and Catherine Wolfram (2001), "Pharmaceutical prices and Political Activity", NBER working paper #8482.

Ferejohn, John (1974), *Pork Barrel Politics; River and Harbors Legislation, 1947-1968*, Stanford University Press.

Finkelstein, Amy (2004), "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry." *Quarterly Journal of Economics,* 119(2), pp. 527-64.

Foucault, Michel (1963) [1994], *The Birth of the Clinic: An Archaeology of Medical Perception*, Vintage.

Griliches, Zvi (1991), "The Search for R&D Spillovers". NBER working paper #3768.

Griliches, Zvi, Tor Jakob Klette and Jarle Møen (2000)"Do Subsidies to Commercial R&D Reduce Market Failures? Microeconomic Evaluation Studies" *Research Policy* Volume 29, Issues 4-5, Pages 471-495.

Grossman, Gene, and Elhanan Helpman (1991), *Innovation and Growth in the Global Economy.* Cambridge, MA: MIT Press.

Hall, Bronwyn H., Jacques Mairesse and Pierre Mohnen (2010), "Measuring the returns to R&D", in *Elsevier Handbook of the Economics of Innovation*, B. H. Hall and N. Rosenberg (eds.), April 2010.

Hayami, Yujiro and Vernon Ruttan (1970), "Factor prices and technical change in agricultural development: the US and japan, 1880-1960," *Journal of Political Economy*, LXXVII, 1115-1141.

Hegde, Deepak (2009), "Political Influence Behind the Veil of Peer Review: An Analysis of Public Biomedical Research Funding in the US". *Journal of Law and Economics* vol. 52 p.665-690.

Jacob, Brian and Lars Lefgren (2007), "The Impact of Research Grant Funding on Scientific Productivity" NBER working paper #13519

Jayachandran, Seema (2007) "The Jeffords Effect", *The Journal of Law and Economics* UChicago Press.

Knight, Brian (2005), "Estimating the Value of Proposal Power", *American Economic Review,* 95(5), 1639-1652.

Krehbiel, Keith (1990), "Are Congressional Committees Composed of Preference Outliers?" *The American Political Science Review,* Vol. 84, No. 1, pp. 149-163.

Levitt, Steven and James Snyder (1997) "The impact of federal spending on house election outcomes." *Journal of Political Economy*, 105.

Lichtenberg, Frank R. (1999), "The Allocation of Publically-Funded Biomedical Research", NBER working paper #6601.

Lichtenberg, Frank R. and Joel Waldfogel (2003) "Does Misery Love Company-Evidence from Pharmaceutical Markets before and after the Orphan Drug Act", NBER Working Paper No. 9750

Londregan, John and James M. Snyder, Jr. (1994), "Comparing Committee and Floor Preferences," *Legislative Studies Quarterly*, Vol. 19, No. 2, pp. 233-266.

Miller, Grant (2008), "Women's Suffrage, Political Responsiveness, and Child Survival in American History," *Quarterly Journal of Economics* 123(3): 1287-1327.

Murphy, Kevin M. and Robert Topel (2003), "The Value of Health and Longevity" *Journal of Political Economy* vol. 114, No. 5.

Payne, Abigail (1999), "Are All Agencies Alike? The Impact of Politics on University Research Funding From Ten Federal Agencies," mimeo.

Payne, Abigail (2003), "The Effects of Congressional Appropriation Committee Membership on the Distribution of Federal Research Funding to Universities," *Economic Inquiry*, Vol. 41, No. 2, 325-345.

Romer, Paul M. (1990), "Endogenous Technological Change," *Journal of Political Economy*, XCVIII, S71-S102.

Samuelson, Paul (1965), "An induced innovation along Kennedy-Weisacker lines", *Review of Economics and Satistics*, XLVII, 444-464.

Schneider, Judy (2007), " House Subcommittees: Assignment Process," *CRS Report for Congress* 98-610 GOV.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz (2006), "Party Influence in Congress and the Economy" NBER Working Paper No. 12751.

Toole, Andrew A. (2007), "Does Public Scientific Research Complement Private Investment in Research and Development in the Pharmaceutical Industry?" *Journal of Law and Economics* vol. 50 p.81-103.

9 Figures and Tables

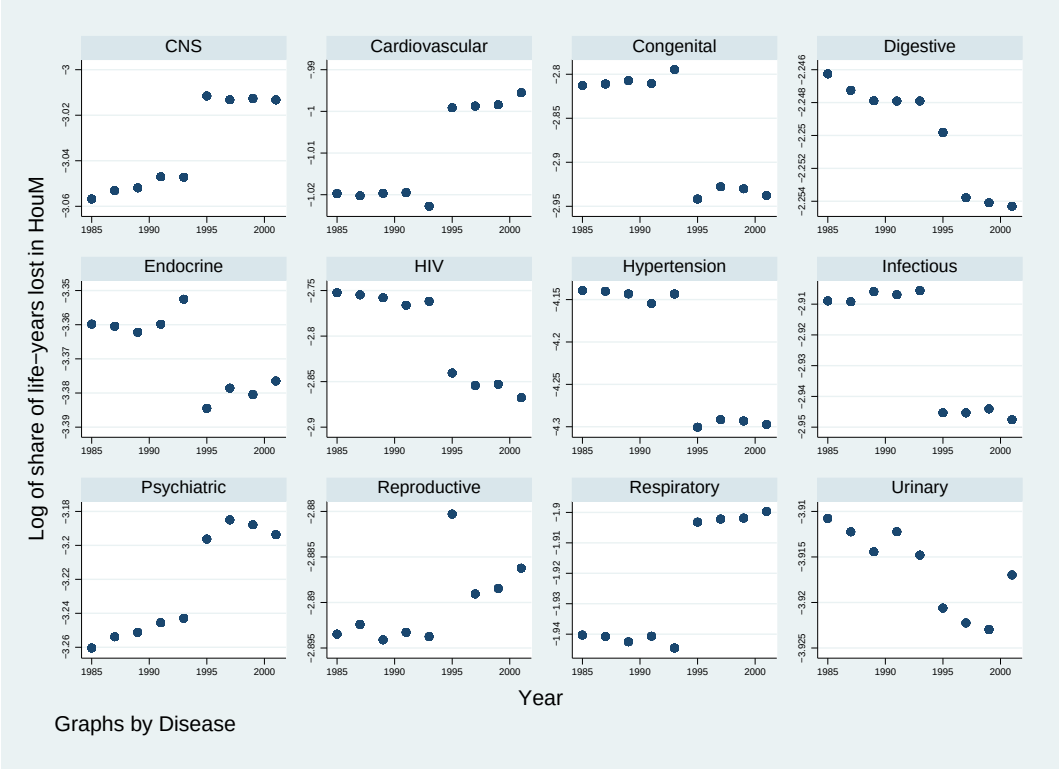

Figure 1: Log of life-years lost in the House majority districts 1985-2002.

Notes: See text for the construction of the data.

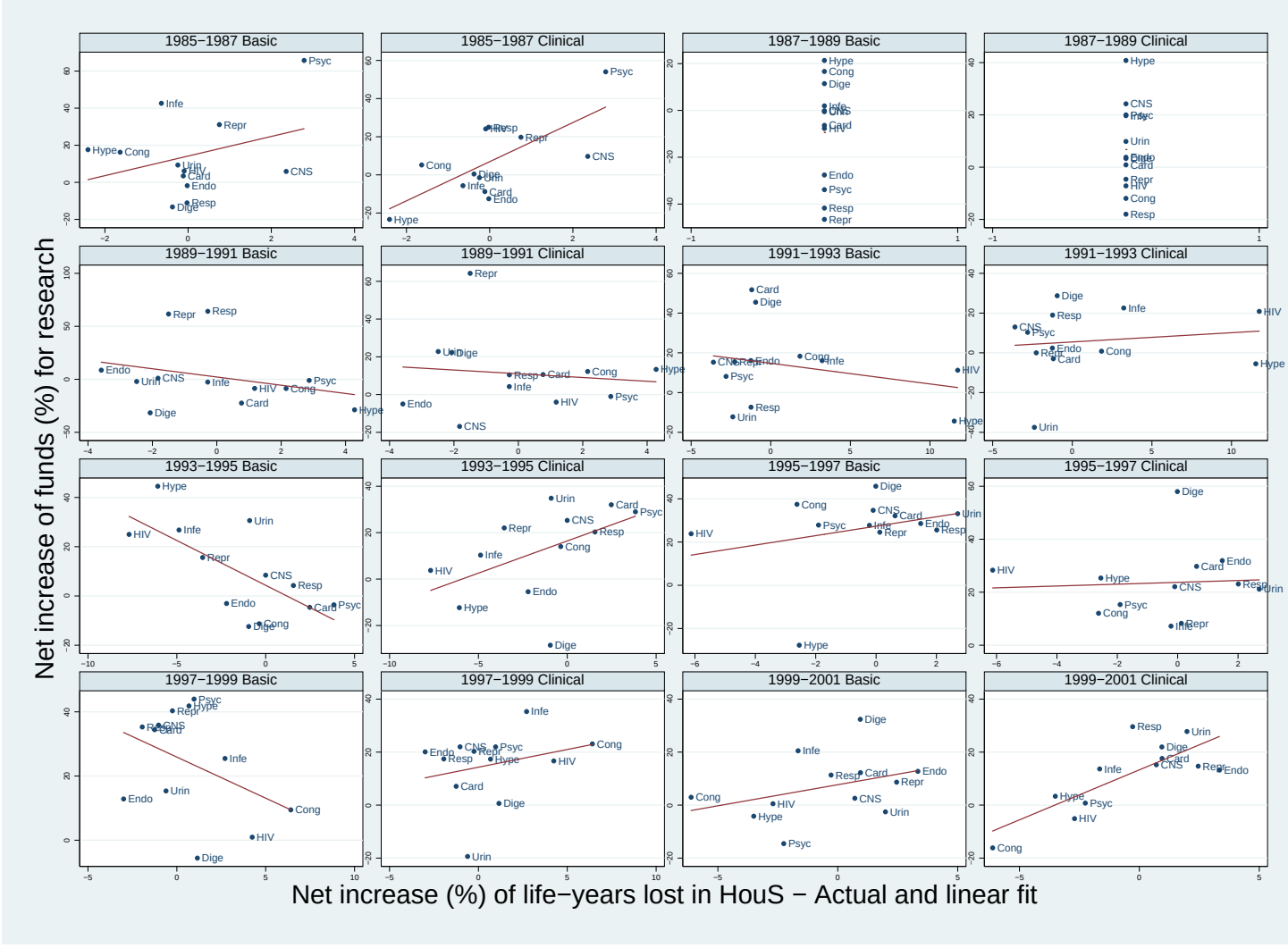

Figure 2: Increase in life-years lost in HouS and increase in funds for research from one Congress term to the next. Actual and linear fit by type of research and by Congress term, for the period 1985-2002.

Notes: See the text for the construction of the data. The x-axis represents the difference between the log of the share of life-years lost in the districts represented in HouS between the current Congress term and the previous one. The full name of diseases is reported in Table 3.

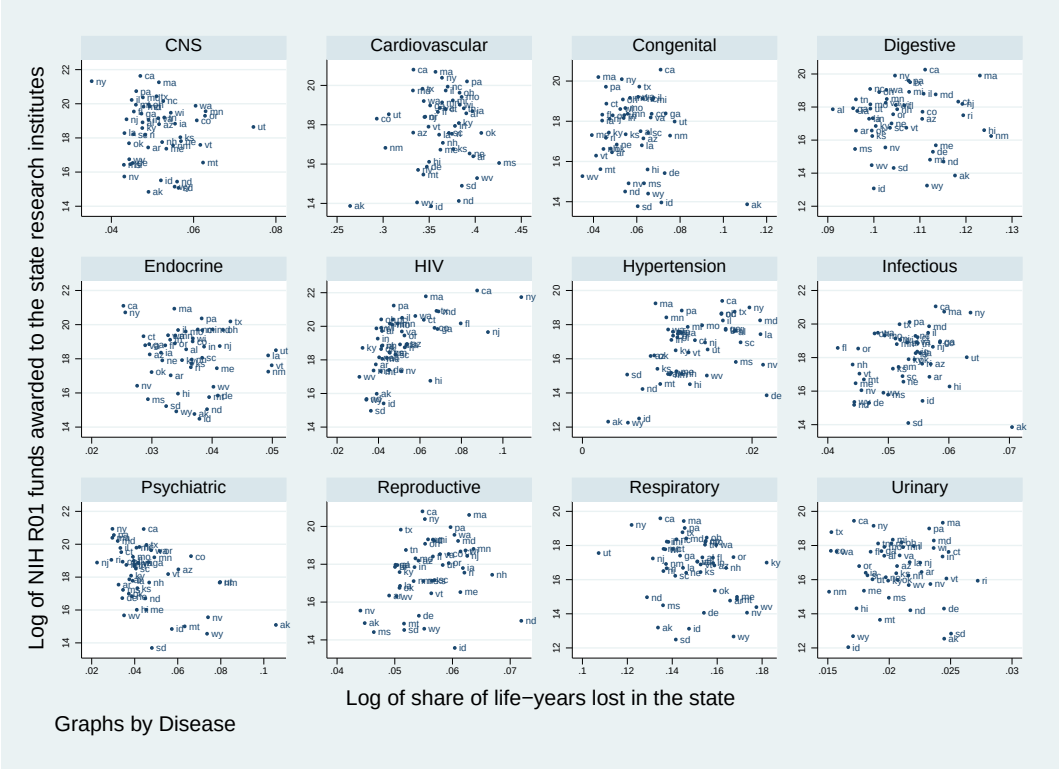

Figure 3: Share of life years lost and NIH funds awarded during the period 1985-2002, by state and by disease.

Notes: See the text for the construction of the data. The x-axis represents the log of the share of life-years lost in any state in 1991.

Table 1: **Congress Groups**

House	Senate	Name/Composition	In charge of:
HouM	SenM	Congressmen who belong to the Majority party of the chamber.	
HouA	SenA	Appropriations committee	the budget proposal
HouS	SenS	Subcommittee on Labor, Health and Human Services, Education and Related Agencies of the Appropriations committees	the NIH budget proposal (among other charges)
HouSM		Representatives who belong both to HouS and to HouM	
HouSm		Representatives who belong to HouS, but not to HouM	
	SenL	Committee on Health, Education, Labor and Pensions	providing recommendations for the NIH budget

Note: The exact name of committees and subcommittees slightly changes from year to year. Section 6.2 describes the construction of the "majority party", a fictitious group that is not included in this table.

Table 2: **Number of Members per Year and Congress Group**

Year	**HouM**	**HouS**	**HouA**	**SenM**	**SenL**	**SenS**	**SenA**
1985	253	13	57	53	16	15	29
1987	258	13	57	55	16	15	29
1989	260	13	57	55	15	15	29
1991	269	12	59	56	17	15	29
1993	259	13	60	55	17	15	29
1995	236	13	56	61	16	15	28
1997	228	14	60	55	17	15	28
1999	223	15	61	55	18	15	28
2001	223	17	65	49	21	15	29

Note: See Table 1.

Table 3: **Summary statistics by disease**

Diseases	**#Deaths (/1000)**	**#Life-years lost(/1000)**	**Log Share of LY lost**			**Log Funds**	
	US	**US**	**US**	**HouM**	**HouS**	**Basic**	**Clinical**
CNS	89.4	2,666	1.57	1.57	1.55	20.4	20.6
				.0205	.0262	.416	.407
Cardiovascular	859	20,158	3.59	3.59	3.6	19.8	19.9
				.0119	.0117	.43	.371
Congenital	34.1	3,185	1.75	1.74	1.7	19.4	19.6
				.0673	.0236	.285	.208
Digestive	197	5,881	2.36	2.36	2.38	19.2	19.4
				.00325	.017	.33	.41
Endocrine	66	1,914	1.24	1.24	1.22	20.1	20.2
				.0116	.0322	.249	.237
HIV	79.6	3,434	1.82	1.8	1.8	20.8	21.2
				.0508	.0436	.255	.296
Hypertension	32.8	830	.403	.394	.448	18.4	18.8
				.08	.0506	.219	.261
Infectious	121	2,992	1.68	1.68	1.68	20.1	19.7
				.0202	.0154	.517	.398
Psychiatric	47.1	2,195	1.37	1.38	1.29	19	20.6
				.0322	.0206	.336	.488
Reproductive	103	3,100	1.72	1.72	1.74	19.1	20
				.00462	.0319	.581	.525
Respiratory	276	8,055	2.67	2.68	2.69	18.6	18.8
				.0212	.0091	.407	.437
Urinary	42.2	1,107	.69	.689	.733	18.6	18.6
				.00459	.022	.315	.231

Notes: columns 1 to 3 report the total number of deaths, life-years lost, and the log of the share of life-years lost (see equation 3) in the US in 1991. Columns 4 and 5 report the mean (upper row) and standard deviation (lower row) of the log of the share of life-years lost in the districts represented in HouM and HouS across Congress terms for the period 1985-2002, by disease. Table 1 contains information on congressional groups. The funds are the NIH funds awarded to R01 grants in the period 1985-2002. CNS stands for *Central Nervous System* diseases, and include connective tissue diseases, musculoskeletal and dermatological disorders. The category *Congenital* disorders includes all death records that are explicitly due to a congenital disease (to compute LY) and any research project (to compute $Log\ Funds$) that mentions *congenital*. The category *HIV* includes all sexually-transmitted disorders, HIV-related pathologies, as well as blood/lymphatic neoplasms (to be consistent with the early NIH classification of HIV). The category *Infectious* diseases includes all infectious disorders except sexually-transmitted disorders. The category *Cardiovascular* diseases includes all cardiovascular disorders that are not explicitly linked to another disorder.

Table 4: **Increase in life-years lost and NIH funds, by disease and type of research**

Congress group G	HouS	HouA	HouM	SenS	SenA	SenM	SenL	HouA	HouM	SenS	SenA	SenM	SenL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Dependent variable is funds by disease and type of research												
Life-years lost in group G	-1.13+	-2.22	-0.76	0.12	-0.46	-0.86	0.24	-1.42	-0.40	0.30	-0.24	-0.75	0.13
	(0.60)	(2.01)	(1.10)	(0.29)	(0.50)	(0.92)	(0.38)	(1.87)	(1.03)	(0.35)	(0.56)	(1.01)	(0.34)
Life-years lost in group G $\times Clinic_\theta$	2.63*	3.67	2.15+	0.16	0.68	0.29	-1.09*	1.53	1.35	-0.23	0.13	0.029	-0.83*
	(0.77)	(2.34)	(1.25)	(0.33)	(0.76)	(1.25)	(0.42)	(2.12)	(1.18)	(0.41)	(0.75)	(1.24)	(0.39)
Life-years lost in HouS								-0.89	-1.02	-1.28+	-1.06	-1.06	-1.10+
								(0.53)	(0.61)	(0.69)	(0.69)	(0.67)	(0.57)
Life-years lost in HouS $\times Clinic$								2.37*	2.28*	2.75*	2.59*	2.63*	2.46*
								(0.75)	(0.84)	(0.89)	(0.88)	(0.84)	(0.78)
Observations	192	192	192	192	192	192	192	192	192	192	192	192	192
R-squared	0.211	0.190	0.195	0.180	0.181	0.186	0.192	0.214	0.218	0.214	0.212	0.220	0.221

Notes: this table reports the OLS estimation of the first difference of equation 12 for the period 1985-2002. The dependent variable *funds by disease and type of research* is the difference between the log of NIH funds for research of type θ on disease d awarded through R01 grants in Congress terms t and $t-1$ ($\log F_{d,\theta,t} - \log F_{d,\theta,t-1}$). *Life-years lost in ...* is the difference between the log of the share of life-years lost to 100 due to disease d in the districts/states represented in ... in Congress terms t and $t-1$ ($\log LY_{d,t} - \log LY_{d,t-1}$). See Table 1 for information on the congressional groups HouS, HouA, etc., and Section 4 for information on the data. *Clinic* is a variable equal to 1 if the type of research is clinical, 0 otherwise. All regressions include dummies for Congress terms and, when relevant, type of research and their interactions. Cluster-robust standard errors are reported in parentheses. * p<0.05, + p<0.1.

Table 5: **Increase in life-years lost and NIH funds, by disease, type of research and state**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable is funds by disease, by type of research and by state							
Life-years lost in HouS	0.46	-0.95	-0.25	-0.25			
	(0.43)	(0.66)	(0.63)	(0.63)			
Life-years lost in HouS $\times Clinic_\theta$		2.83*	1.85*	1.85*			
		(0.86)	(0.84)	(0.84)			
Life-years lost in HouA			-5.06*	-5.05*	-3.74*		-5.05*
			(1.77)	(1.77)	(1.49)		(1.78)
Life-years lost in HouA $\times Clinic_\theta$			1.98	1.97	5.04*		3.00
			(2.31)	(2.31)	(1.92)		(2.30)
Life-years lost in HouM			0.93	0.93		-0.36	0.91
			(0.75)	(0.75)		(0.75)	(0.81)
Life-years lost in HouM $\times Clinic_\theta$			1.40	1.40		2.48*	1.97+
			(1.07)	(1.07)		(1.00)	(1.12)
Life-years lost in SenS			-0.64	-0.64			
			(0.42)	(0.43)			
Life-years lost in SenS $\times Clinic_\theta$			0.12	0.12			
			(0.57)	(0.57)			
Life-years lost in SenA			1.18	1.18			0.63
			(0.81)	(0.81)			(0.68)
Life-years lost in SenA $\times Clinic_\theta$			0.36	0.36			0.79
			(1.13)	(1.14)			(0.92)
Life-years lost in SenM			-2.27*	-2.27*			-1.75*
			(0.86)	(0.86)			(0.80)
Life-years lost in SenM $\times Clinic_\theta$			-0.69	-0.68			-0.83
			(1.24)	(1.24)			(1.13)
Life-years lost in SenL			-0.44	-0.44			
			(0.39)	(0.39)			
Life-years lost in SenL $\times Clinic_\theta$			-0.90+	-0.90+			
			(0.52)	(0.52)			
# State congressmen variables	No	No	No	Yes	No	No	No
Observations	9,475	9,475	9,475	9,475	9,475	9,475	9,475
R^2	0.017	0.018	0.022	0.025	0.018	0.018	0.020

Notes: this table reports the OLS estimation of the first difference of equation 13 for the period 1985-2002. The dependent variable *funds by disease, type of research and state* is the difference between the log of NIH funds for research of type θ on disease d awarded through R01 grants to research institutes in state s in Congress terms t and $t-1$ ($\log F_{d,\theta,s,t} - \log F_{d,\theta,s,t-1}$). *Life-years lost in ...* is the difference between the log of the share of life-years lost to 100 due to disease d in the districts/states represented in ... in Congress terms t and $t-1$ ($\log LY_{d,t} - \log LY_{d,t-1}$). Column (4) includes controls for the number of congressmen of state s represented in any group ($NC_{s,t} - NC_{s,t-1}$). See Table 1 for information on the groups, Section 4 for information on the data. *Clinic* equals 1 if the type of research is clinical, 0 otherwise. All regressions include dummies for Congress terms and type of research and their interactions. Cluster-robust standard errors are in parentheses. * p<0.05, + p<0.1.

Table 6: **Increase in life-years lost and NIH funds - in levels**

	(1)	(2)	(3)	(4)
	Dependent variable is funds by disease and by type of research			
		and by state		
Life-years lost in HouS	-31,859	-640*	-591*	-1,213*
	(34,178)	(299)	(274)	(272)
Life-years lost in HouS $\times Clinic_\theta$	102,280*	2,050*	1,827*	2,607*
	(40,188)	(513)	(478)	(531)
Life-years lost in HouA				589
				(650)
Life-years lost in HouA $\times Clinic_\theta$				-101
				(1,149)
Life-years lost in HouM			-393	-172
			(449)	(470)
Life-years lost in HouM $\times Clinic_\theta$			1,786*	1,288
			(722)	(751)
Life-years lost in SenS				6.29
				(9.87)
Life-years lost in SenS $\times Clinic_\theta$				-17.7
				(21.0)
Life-years lost in SenA				23.9*
				(7.84)
Life-years lost in SenA $\times Clinic_\theta$				-30.5*
				(13.9)
Life-years lost in SenM				-25.2*
				(8.99)
Life-years lost in SenM $\times Clinic_\theta$				28.0
				(14.3)
Life-years lost in SenL				6.88
				(4.27)
Life-years lost in SenL $\times Clinic_\theta$				-4.69
				(8.91)
Observations	192	9,534	9,534	9,534
R^2	0.243	0.026	0.027	0.029

Notes: this table reports the OLS estimation of the first difference of equation 12 in Column 1 (resp. equation 13 in Columns 2 to 4) for the period 1985-2002 with the following modifications. The dependent variable *funds by disease, type of research* (resp. *funds by disease, type of research, and state*) is the difference between the NIH funds awarded for research of type θ, on disease d (resp. of type θ, on disease d, to research institutes in state s) in Congress terms t and $t-1$ ($F_{d,\theta,t} - F_{d,\theta,t-1}$ (resp. $F_{d,\theta,s,t} - F_{d,\theta,s,t-1}$)). *Life-years lost in ...* is the difference between the average number of life-years lost to 100 due to disease d in the districts/states represented in ... in Congress terms t and $t-1$ ($\ell y_{d,t} - \ell y_{d,t-1}$, see equation 4). See Table 1 for information on the groups. *Clinic* equals 1 if the type of research is clinical, 0 otherwise. All regressions include dummies for Congress terms and type of research and their interactions. Cluster-robust standard errors are in parentheses. * p<0.05, + p<0.1.

Table 7: **Increase in life-years lost on NIH funds - 2SLS**

	OLS	2SLS	2SLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)
Dependent variable	Life-years lost in HouS	funds by disease and type of research		funds by disease, type of research, and state	
Life-years lost in HouS		-1.66	-1.34	0.19	0.97
		(1.07)	(1.11)	(1.51)	(1.53)
Life-years lost in HouS $\times Clinic_\theta$		3.12*	1.19	2.59	-0.27
		(1.30)	(1.73)	(1.92)	(1.90)
Life-years lost in HouM			-0.29		-0.70
			(1.06)		(0.69)
Life-years lost in HouM $\times Clinic_\theta$			1.73		2.57*
			(1.37)		(0.92)
Life-years lost in HouS*	3.17*				
	(0.32)				
Observations	192	192	192	9,475	9,475
R^2	0.323	0.209	0.200	0.018	0.018

Notes: this table reports the 2SLS estimation of the parameters of equation 12 in Columns 2-3 (resp. equation 13 in Columns 4-5) for the period 1985-2002, using Life-years lost in HouS* as an instrument for Life-years lost in HouS (see equation 15). The dependent variable *funds by disease, type of research* (resp. *and state*) is the difference between the NIH funds for research of type θ, on disease d, awarded through R01 grants (resp. to research institutes in state s), in Congress terms t and $t-1$ ($\log F_{d,\theta,t} - \log F_{d,\theta,t-1}$ (resp. $\log F_{d,\theta,s,t} - \log F_{d,\theta,s,t-1}$)). *Life-years lost in ...* is the difference between the log of the share of life-years lost to 100 due to disease d in the districts/states represented in ... in Congress terms t and $t-1$ ($LY_{d,t} - LY_{d,t-1}$). Column 1 reports the estimation of the first stage. Table 1 provides information on congressional groups. $Clinic_\theta$ is a variable equal to 1 if $\theta = clinical$, 0 otherwise. The regressions include dummies for Congress terms and, when relevant, type of research and their interactions. * p<0.05, + p<0.1.

Table 8: **Role of House majority**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable is funds by disease, by type of research						and by state	
Life-years lost in HouSM	-0.84*	-1.19*				-1.18*	
	(0.38)	(0.29)				(0.44)	
Life-years lost in HouSM $\times Clinic_\theta$	1.85*	1.43*				1.42*	
	(0.48)	(0.54)				(0.62)	
Life-years lost in HouSm	-0.027	0.22				0.32	
	(0.46)	(0.58)				(0.52)	
Life-years lost in HouSm $\times Clinic_\theta$	0.54	0.84				0.93	
	(0.56)	(0.70)				(0.67)	
Life-years lost in HouM		1.14		5.17		1.66	
		(1.59)		(3.86)		(1.36)	
Life-years lost in HouM $\times Clinic_\theta$		1.41		-2.85		1.87	
		(1.89)		(5.11)		(1.83)	
Life-years lost in majority party			-1.08	-6.29+	-0.37		-0.67
			(0.96)	(3.21)	(0.98)		(0.71)
Life-years lost in maj. party $\times Clinic_\theta$			2.43*	5.31	2.48+		2.52*
			(1.08)	(4.45)	(1.30)		(0.95)
Life-years lost in HouA					-2.40		
					(2.00)		
Life-years lost in HouA $\times Clinic_\theta$					2.18		
					(2.34)		
Life-years lost in SenA					0.35		
					(0.60)		
Life-years lost in SenA $\times Clinic_\theta$					-0.092		
					(0.84)		
Life-years lost in SenM					-0.96		
					(1.17)		
Life-years lost in SenM $\times Clinic_\theta$					-0.68		
					(1.53)		
Observations	192	192	192	192	192	9,475	9,475
R^2	0.218	0.234	0.197	0.208	0.221	0.020	0.018

Notes: this table reports the OLS estimation of the first difference of equation 12 in columns 1-5 (resp. equation 13 in columns 6-7) for the period 1985-2002. The dependent variable *funds by disease and by type of research* (resp. *by disease, by type of research and by state*) is the difference between the log of NIH funds for research of type θ on disease d awarded (resp. of type θ on disease d awarded to research institutes in state s) through R01 grants in Congress terms t and $t-1$ ($\log F_{d,\theta,t} - \log F_{d,\theta,t-1}$ [resp. $\log F_{d,\theta,s,t} - \log F_{d,\theta,s,t-1}$]). *Life-years lost in ...* is the difference between the log of the share of life-years lost to 100 due to disease d in the districts/states represented in ... in Congress terms t and $t-1$ ($\log LY_{d,t} - \log LY_{d,t-1}$). See Table 1 for information on the groups, Section 6.2 describes the construction of the fictitious "majority party". *Clinic* equals 1 if the type of research is clinical, 0 otherwise. All regressions include dummies for Congress terms and type of research and their interactions. Cluster-robust standard errors are in parentheses. * p<0.05, + p<0.1.

Table 9: **Estimation of the effect of 1995 majority shift**

	All		*Clinical*	
Disease	**Actual**	**Counterfactual**	**Actual**	**Counterfactual**
CNS	9.71e+09	9.63e+09	5.25e+09	4.99e+09
Cardiovascular	5.13e+09	5.11e+09	2.69e+09	2.60e+09
Congenital	3.07e+09	3.20e+09	1.63e+09	1.96e+09
Digestive	2.63e+09	2.63e+09	1.49e+09	1.50e+09
Endocrine	5.51e+09	5.52e+09	2.75e+09	2.82e+09
HIV	1.42e+10	1.49e+10	8.76e+09	9.99e+09
Hypertension	1.21e+09	1.30e+09	7.32e+08	8.99e+08
Infectious	5.69e+09	5.66e+09	2.17e+09	2.29e+09
Psychiatric	6.38e+09	6.04e+09	5.45e+09	5.06e+09
Reproductive	4.48e+09	4.45e+09	3.08e+09	3.04e+09
Respiratory	1.57e+09	1.56e+09	8.99e+08	8.52e+08
Urinary	1.27e+09	1.27e+09	5.96e+08	6.01e+08

Notes: the table presents the estimation of the actual and counterfactual total amount of funds that would have been awarded through competing R01 grants between 1995 and 2002 for any type of research (*All*), and for clinical research only (*Clinical*), if the House majority had not changed in 1995. The counterfactual estimations rely on the specification of column (3) in Table 8.

Table 10: **Effect of state affiliation of congressmen on NIH funds**

	(1)	(2)	(3)	(4)	(5)
Dependent variable is funds by disease, by type of research and by state					
# State congressmen in HouS	-0.048*	-0.030	-0.031	0.0064	0.0064
	(0.024)	(0.034)	(0.034)	(0.037)	(0.037)
# State congressmen in HouS $\times Clinic_\theta$		-0.036	-0.033	-0.047	-0.047
		(0.048)	(0.048)	(0.054)	(0.054)
# State congressmen in HouA				-0.058+	-0.058+
				(0.031)	(0.031)
# State congressmen in HouA $\times Clinic_\theta$				0.015	0.015
				(0.040)	(0.040)
# State congressmen in HouM			0.0064	0.0081	0.0081
			(0.0095)	(0.0096)	(0.0096)
# State congressmen in HouM $\times Clinic_\theta$			-0.013	-0.016	-0.016
			(0.013)	(0.013)	(0.013)
# State congressmen in SenS				0.098	0.097
				(0.064)	(0.063)
# State congressmen in SenS $\times Clinic_\theta$				0.042	0.043
				(0.088)	(0.088)
# State congressmen in SenA				-0.094+	-0.094+
				(0.049)	(0.049)
# State congressmen in SenA $\times Clinic_\theta$				0.018	0.018
				(0.065)	(0.065)
# State congressmen in SenM				-0.026	-0.026
				(0.021)	(0.021)
# State congressmen in SenM $\times Clinic_\theta$				0.047	0.048
				(0.030)	(0.030)
# State congressmen in SenL				0.030	0.030
				(0.044)	(0.045)
# State congressmen in SenL $\times Clinic_\theta$				0.044	0.044
				(0.058)	(0.058)
Life-years lost variables	No	No	No	No	Yes
Observations	9,475	9,475	9,475	9,475	9,475
R^2	0.017	0.017	0.017	0.019	0.025

Notes: this table reports the OLS estimation of the first difference of equation 13 for the period 1985-2002. The dependent variable *funds by disease, type of research and state* is the difference between the log of NIH funds for research of type θ on disease d awarded through R01 grants to research institutes in state s in Congress terms t and $t-1$ ($\log F_{d,\theta,s,t} - \log F_{d,\theta,s,t-1}$). *# State congressmen in ...* is the difference between the number of congressmen from state s in ... in Congress terms t and $t-1$ ($NC_{s,t} - NC_{s,t-1}$). Column (5) includes controls for the life-years lost in any group ($\log LY_{d,t} - \log LY_{s,t-1}$). See Table 1 for information on the groups, Section 4 for information on the data. *Clinic* equals 1 if the type of research is clinical, 0 otherwise. All regressions include dummies for Congress terms and type of research and their interactions. Cluster-robust standard errors are in parentheses. * p<0.05, + p<0.1.

Table 11: **Effect of local life-years lost on research focus**

	(1)	(2)	(3)
Dependent variable is funds by disease, type of research and state			
Life-years lost in state s	0.087	0.084	0.14
	(0.088)	(0.073)	(0.12)
Life-years lost in state s $\times Clinic_\theta$			-0.12
			(0.14)
Interaction Congress term-$Clinic_\theta$	No	Yes	Yes
Interaction Disease-$Clinic_\theta$	No	Yes	Yes
Observations	10,729	10,729	10,729
R^2	0.849	0.863	0.863

Notes: this table reports the OLS estimation of equation 18 for the period 1985-2002. The dependent variable *funds by disease, type of research and state* is the log of NIH funds for research of type θ on disease d awarded through R01 grants to research institutes in state s in Congress terms t and $t-1$ ($\log F_{d,\theta,s,t}$). *Life-years lost in state s* is the log of the share of life-years lost to 100 due to disease d in the districts/states represented in state s in Congress terms t ($\log LY^s_{d,t}$). *Clinic* equals 1 if the type of research is clinical, 0 otherwise. All regressions include dummies for Congress terms, diseases, types of research and states. Cluster-robust standard errors in parentheses. * p<0.05, + p<0.1.

Table 12: **Variance of funds for basic and clinical research**

	(1)	(2)	(3)	(4)	(5)
Dependent variable is the square of the predicted residual $\epsilon^2_{d,\theta,t}$ of the estimation reported in Column (9) of Table 4					
$Clinic_\theta$	-0.014	-0.014	-0.014*	-0.014*	-0.014*
	(0.0092)	(0.0094)	(0.0047)	(0.0047)	(0.0064)
Congress term t dummy	No	Yes	No	Yes	Yes
Disease d dummy	No	No	Yes	yes	Yes
Interaction Disease-Congress Term	No	No	No	No	Yes
Observations	192	192	192	192	192
R^2	0.014	0.071	0.120	0.177	0.711

Notes: this table reports OLS estimation of the impact of the type of research (basic or clinical) on the squared predicted residual of the estimation presented in Column (9) in Table 4. $Clinic_\theta$ is a variable equal to 1 if $\theta = clinical$, 0 otherwise. Cluster-robust standard errors in parentheses. * p<0.05, + p<0.1.

10 Appendix

Proof of Proposition 1.

Consider the solution $\left(L^*, \mathbf{F}^*_{L^*}\right)$ defined in the proposition. If $2 \in L^*$, then taking $\left(\tilde{L}^*, \mathbf{F}^*_{\tilde{L}^*}\right) = \left(L^*, \mathbf{F}^*_{L^*}\right)$ proves the claim. If $2 \notin L^*$, let the function $V : [0,1] \to \mathbb{R}$ be defined as:

$$\begin{aligned} V(\delta) \quad = \quad & \int_0^{+\infty} \dots \int_0^{+\infty} \left((1-\delta) c_\alpha A_1^C {F_1^*}^{\gamma} + c_\alpha \delta A_1^C \hat{F}_\alpha^\gamma \right) g_1(c_\alpha) dc_\alpha g_2(c_2) dc_2 g_3(c_3) dc_3 \dots g_n(c_n) dc_n + \\ & \int_0^{+\infty} \dots \int_0^{+\infty} \left(\delta c_\alpha A_2^C {F_1^*}^{\gamma} + (1-\delta) c_\alpha A_2^C \hat{F}_\alpha^\gamma \right) g_1(c_1) dc_1 g_2(c\alpha) dc\alpha g_3(c_3) dc_3 \dots g_n(c_n) dc_n + \\ & \sum_{k \in L^* \setminus \{1\}} \int_0^{+\infty} c_k {F_k^*}^{\gamma} A_k^C g_k(c_k) dc_k + \\ & \sum_{k \in K \setminus L^* \cup \{2\}} \int_0^{+\infty} \dots \int_0^{+\infty} c_k \hat{F}_k^\gamma A_k^C g_1(c_1) dc_1 g_2(c_2) dc_2 \dots g_n(c_n) dc_n \end{aligned}$$

where $\hat{F}_\alpha, \hat{\mathbf{F}}_{K \setminus (L^* \cup \{2\})}$ solves:

$$\begin{cases} Max_{\{F_\alpha \in \mathbb{R}^+, F_k \in \mathbb{R}^+, k \in K \setminus (L^* \cup \{2\})\}} c_\alpha F_\alpha^\gamma A_1^N + \sum_{k \in K \setminus (L^* \cup \{2\})} c_k F_k^\gamma A_k^N \\ s.t.\ F_\alpha + \sum_{K \setminus (L^* \cup \{2\})} F_k = F^* - \sum_{k \in L^*} F_k^* \end{cases} \tag{19}$$

$V(0)$ is C's utility achieved at the solution $\left(F^*, L^*, \mathbf{F}^*_{L^*}\right)$, and $V(1)$ is the utility that would be achieved if 1 and 2 were swapped, every other variable chosen by C remaining unchanged (note that the expected allocation of funds across the research categories of $K \setminus (L^* \cup \{2\})$ by N is then the same before and after the swap). The first-order conditions of N's problem impose, for any $k \in \{\alpha\} \cup K \setminus (L^* \cup \{2\})$:

$$\hat{F}_k = \frac{(c_k A_k)^{\frac{1}{1-\gamma}}}{\sum_{k \in \{c\} \cup K \setminus (L^* \cup \{2\})} (c_k A_k)^{\frac{1}{1-\gamma}}} \left(F^* - \sum_{k \in L^*} F_k^*\right)$$

so that:

$$V'(\delta) = A_1^C \int_0^{+\infty} \dots \int_0^{+\infty} c_\alpha \hat{F}_c^\gamma [g_1(c_\alpha) - g_2(c_\alpha)] dc_\alpha \Pi_{k \in K \setminus (L^* \cup \{2\})} g_k(c_k) dc_k$$

is nonnegative since the function $c_\alpha \mapsto c_\alpha \hat{F}_\alpha^\gamma$ is convex and g_2 second-order dominates g_1. V is nondecreasing in δ, $V(0) \leq V(1)$, so that the solution defined by the swap of 1 and 2 is indeed an optimum of C's maximization problem, in which 2 is not delegated.

Table 13: **NIH Funds awarded through R01 grants- Total and sample**

Year	Total funds	Sample funds	Sample - Basic	Sample - Clinical
1985	1.02e+10	7.43e+09	3.30e+09	4.13e+09
1987	1.06e+10	8.37e+09	3.68e+09	4.69e+09
1989	9.89e+09	8.37e+09	3.41e+09	4.96e+09
1991	9.89e+09	8.52e+09	3.34e+09	5.18e+09
1993	1.14e+10	9.71e+09	3.95e+09	5.76e+09
1995	1.29e+10	1.10e+10	4.41e+09	6.59e+09
1997	1.68e+10	1.41e+10	5.87e+09	8.28e+09
1999	2.03e+10	1.72e+10	7.23e+09	9.94e+09
2001	2.19e+10	1.85e+10	7.82e+09	1.07e+10

Notes: total of NIH funds awarded through competing R01 grants, and sample of funds that I can match to a disease and a type of research, by Congress term, for the period 1985-2002.

Table 14: **Summary statistics**

Variable	Mean	Std. Dev.	Min.	Max.
Funds $F_{d,s,\theta,t}$	9619018.177	21966644.853	77	402696000
Log $F_{d,s,\theta,t}$	14.411	2.253	4.353	19.814
# State congressmen in HouS	0.275	0.516	0	2
# State congressmen in HouA	1.19	1.354	0	7
# State congressmen in HouM	4.935	5.139	0	30
# State congressmen in SenS	0.299	0.482	0	2
# State congressmen in SenA	0.573	0.55	0	2
# State congressmen in SenM	1.089	0.771	0	2
# State congressmen in SenL	0.34	0.497	0	2
# State congressmen in majority and HouS	0.17	0.41	0	2
# State congressmen in minority and HouS	0.105	0.314	0	2
# State congressmen in majority party	5.116	5.582	0	31.909
$LY_{d,t}$ in subcommittee HouS ($\times 100$)	8.343	9.220	1.478	37.202
$LY_{d,t}$ in committee HouA ($\times 100$)	8.343	9.197	1.433	36.734
$LY_{d,t}$ in House majority ($\times 100$)	8.343	9.152	1.356	36.953
$LY_{d,t}$ in subcommittee SenS ($\times 100$)	8.343	9.355	1.374	37.648
$LY_{d,t}$ in committee SenA ($\times 100$)	8.343	9.130	1.472	37.064
$LY_{d,t}$ in Senate majority ($\times 100$)	8.343	9.19	1.404	37.071
$LY_{d,t}$ in committee SenL ($\times 100$)	8.343	9.332	1.294	37.112
$LY_{d,t}$ in House majority and HouS	8.343	9.309	1.377	38.274
$LY_{d,t}$ in House minority and HouS	8.343	9.124	1.246	37.335
$LY_{d,t}$ in House majority party	8.343	9.128	1.365	36.828
Life-years lost in HouS	-2.869	0.815	-4.215	-0.989
Life-years lost in HouA	-2.869	0.820	-4.245	-1.001
Life-years lost in HouM	-2.867	0.819	-4.3	-0.996
Life-years lost in SenS	-2.879	0.830	-4.287	-0.977
Life-years lost in SenA	-2.864	0.816	-4.218	-0.993
Life-years lost in SenM	-2.868	0.818	-4.266	-0.992
Life-years lost in SenL	-2.876	0.825	-4.347	-0.991
Life-years lost in majority and HouS	-2.875	0.821	-4.286	-0.96
Life-years lost in minority and HouS	-2.864	0.814	-4.385	-0.985
Life-years lost in majority party	-2.865	0.818	-4.294	-0.999

Notes: $F_{d,\theta,s,t}$ denotes funds for research on disease d of type θ, awarded to research institutes in state s in Congress term t. $LY_{d,t}$ in group G: share of life-years lost to 100 due to disease d in the districts/states represented in the group G at t as defined in equation 3. Table 1 contains information on congressional groups. "Majority party" is a constructed group defined in Section 6.2