Tim Jaekel

PEER REVIEW IN PUBLIC SECTOR ORGANIZATIONS: A GENERAL MODEL AND EMPIRICAL EVIDENCE FROM A SURVIVAL ANALYSIS

BASIC RESEARCH PROGRAM

WORKING PAPERS

Tim Jaekel[1]

PEER REVIEW IN PUBLIC SECTOR ORGANIZATIONS: A GENERAL MODEL AND EMPIRICAL EVIDENCE FROM A SURVIVAL ANALYSIS[2]

In my paper I analyze why some top-level public administrators invite a peer review to learn about the strengths and weaknesses of their agencies while others do not. A peer review is a light-touch voluntary benchmarking exercise conducted by a group of critical friends (peers). I propose a general model from which I derive a series of hypotheses about the role of organizational size, performance gaps, peer effects and strategic interaction at individual and organizational-level decision making. For hypotheses tests I examine a unique dataset of participation in the Corporate Peer Challenge Program in England between 2010 and 2015. The estimation approach is survival analysis. I find that poor archival performance of a council and peer evaluations in neighboring councils are positively correlated with inviting a peer review. However, significance level of both effects is above 10 per cent.

JEL Classification Codes: D72, D73, D81

Keywords: Public Administration, decision making, performance evaluation, peer review, England, corporate peer challenge program, event history analysis, survival analysis

---

[1] National Research University Higher School of Economics: School of Public Administration; Assistant Professor. 20 Myasnitskaya Ulitsa, Moscow, Russian Federation, 101000. E-mail: tjekel@hse.ru

Basic idea and research question

Decision makers in public administration repeatedly have to make a choice between exploration of new opportunities and exploitation of existing knowledge (March and Simon 1993 [1958], March 1991, Choi and Chandler 2015). Heads of paid service in government agencies are facing a trade-off between being innovative by trying "new ways of doing things" (Fernandez and Moldogaziev 2013), and using existing knowledge stored in the organizational memory of their bureaucracy. Both investment strategies, search for new knowledge and use of existing knowledge given in an organization's internal repository (March and Simon 1993 [1958]), are essential for survive-oriented organizations, but compete for scarce personnel and fiscal resources (March 1991: p. 71; Choi and Chandler 2015). An optimal investment strategy will be the outcome according to rational theories of decision making (Edwards 1954); while theories of limited rationality (Gigerenzer and Goldstein 1996) state that a satisficing resource allocation decision is likely to occur (Simon 1997 [1948]; March and Simon 1993 [1958]).

Inviting a peer-reviewed based performance evaluation provides an excellent example for this strategic trade-off. It will thus form the context of the following analysis. What is a peer-review base performance evaluation? It is not an inspection, because it is non-compulsory, but voluntary. It is not carried out by a government watchdog and its inspectors, but by trustworthy peers, that is, administrative professionals from other agencies. These peers are often denoted as critical friends. The phenomena I am writing about is a voluntary review of an organization's strengths and weaknesses, both in terms of policy performance but even more about managerial capacity. Such non-compulsory light-touch benchmarking exercises have been reported both for local governments and their bureaucracies in the US (Ammons and Roenigk 2015, Ammons and Rivenbark 2008), Australia and European countries, including Germany, the Scandinavian countries, England, or Switzerland (Kuhlmann and Jaekel 2013, Jaekel 2016).

Why do councils invite such a non-compulsory peer review? Sometimes the story behind is straightforward. Bedford council, a local council in the East of England, in 2009 experienced two routine inspections from Ofsted, the English school and social service inspectorate. The council's children services were rated at level 2 and 3 which is somewhere between weak and fair performance; but the inspections induced no priority actions. The big surprise came when in January 2012 a full scale Ofsted inspection judged the children services as inadequate. The

council needed help, and a peer review was supported by the Mayor and the chief executive. The review was undertaken by an ex-director of children services. The chief executive later said that the review provided him with an external focus across all services, also beyond the focus of the Ofsted watchdog. The Bedford story is the textbook story. The ingredients are a performance gap revealed by a mandatory inspection, in this case Ofsted, and a need to act, plus a mayor and a chief executive supporting the idea; and finally a council approving it.
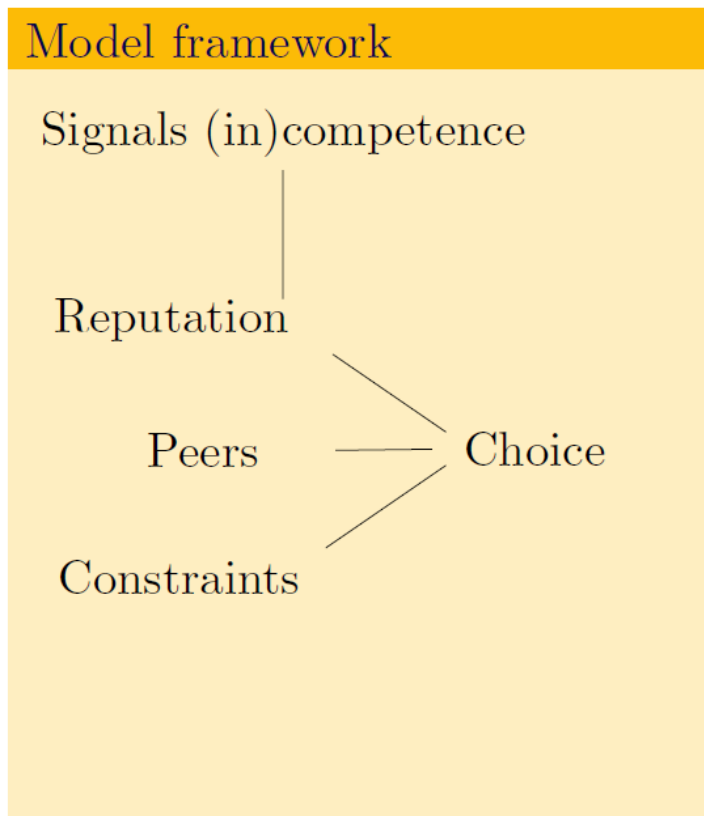
My general research question is why some administrative professionals and their government agencies invite peer reviews while others do not. This question is at the core of organizational and administrative sciences. Innovative behavior in general has been found to be correlated with an array of factors, including professionals hired from outside (Teodoro 2011), empowerment practices (Fernandez and Moldogaziev 2013), diffusion spillover among peers (Arnold 2014), but also 'clear signals' from above, that is, expectations from central government agencies (Zhu and Zhang 2015). However, empirical evidence on the adaption of peer evaluations is rare (Jaekel 2016), mostly based on qualitative case studies and thus mixed.

A general model of peer review participation

I develop a general model in which the decision of an administrative professional and her agency whether or not to participate in a non-compulsory performance evaluation results from

- A relative performance gap,
- Spillover among nearby councils
- Peer effects among professionals that share interpersonal and/or inter-organizational ties.
- Fundamental agency level characteristics including size (constraints)

Figure 1: Model framework

Note: Own figure.

Constraints

In addition, running a bureaucracy is mainly dealing with constraints; professionals to a certain degree have to satisfy the expectations of external and internal stakeholders (Wilson 1989). The observed outcome will be correlated to an array of organizational features, including size. Early studies on innovation found that organizational attributes have no or only limited effect on innovative behavior (Wilson 1989). More recent studies find that size and some other features do affect innovative behavior: Large municipalities are more likely to invite a Peer Review. They have more leeway to invest some time and staff.

> *Hypothesis 2 (Size):* The size of an organization will be positively related to invitation of a peer peview.

Performance and Reputation

For a senior decision maker inviting a team of critical friends to her agency is a potential mean for receiving a fair judgment on her level of competence and areas for improvement in

service delivery activities. She might expect two main benefits: first a virtual certificate of her levels of competence; and second suggestions how to do better. Both will help her to build, maintain or improve her professional reputation and thus job market prospects.

On the one hand underperforming organizations should have a strong incentive to learn and to improve. On the other hand the model considers two mechanisms that work against this: First showcasing underperformance scratches performance reputation and thus job market prospects. A review that reveals incompetence will scratch your reputation among peers. Second there is also a psychological barrier. Robert Kramer argues that the hardest thing for chief executives is to admit that they do not know how to solve an adaptive problem (Kramer 2015, under review). In hierarchic bureaucracy civil servants have strong incentives to hide failures.

This situation can be formulated as an expected utility model: Subjects choose among two risky outcomes. I argue that inviting a Peer Review is a decision to explore new opportunities to learn about areas of improvement. Not to invite a peer review can be understood as a choice for the exploitation of existing knowledge. The subject will evaluate the prospect X of each choice; which equals the sum of the utilities weighted by the probabilities attached to each outcome.

$$X = (p_1 x_1; p_2 x_2)$$

$$X \rightarrow \sum p_i \, u \, (x_i),$$

where x_1 is gains, and x_2 is losses, both in terms of performance reputation. P_i is a proxy for prior performance, in whatever terms.

Inviting a peer review yields either a gain (in terms of performance reputation), or a loss. But what is the probability attached to each outcome? I argue that performance measurements from prior performance assessments, archival performance scores, are a hint. Examples for archival performance scores include a performance rating from a previous service inspection, the level of customer satisfaction reveled by a recent citizens survey, generally any value of a policy-related key performance indicator. A low archival performance score is associated with a high probability of a loss in terms of performance reputation. A high archival performance score is associated with a high probability of a gain in terms of performance reputation.

My expected utility model states that the expected reputation improvement is weighted against the expected reputational costs. The relationship between performance and inviting a

peer review could be positive or negative. It could also be that the incentives to improve and the incentives to hide underperformance cancel out each other.

> *Hypothesis 1a:* Organizations with *low* performance (in terms of standardized performance scores) are *less* likely to invite a Corporate Peer Challenge (CPC).

> *Hypothesis 1b:* Organizations with low performance (in terms of standardized performance scores) are more likely to invite a Corporate Peer Challenge (CPC).

Note that Hypothesis 1a and 1b are contradictory to each other. From a theoretical perspective both can be derived.

Peer effects

Administrative professional do not act in a closed environment. Administrators intend to comply with professional norms to maximize their future job prospects (Teodoro 2011). They learn from successful peers by copying their behavior (Arnold 2015, Huang 2014). On organizational level mimicry results in mimetic isomorphism (DiMaggio and Powell 1983). The model states that the decision whether to invite a peer review will be affected by the decisions of relevant peers, and nearby councils.

The model distinguishes between two different mechanisms: In the former case, spillover results from personal linkages between senior professionals. In the latter case councils strategically interact because they compete for mobile taxpayers, government resources and investments (Besley and Case 1995; yardstick competition).
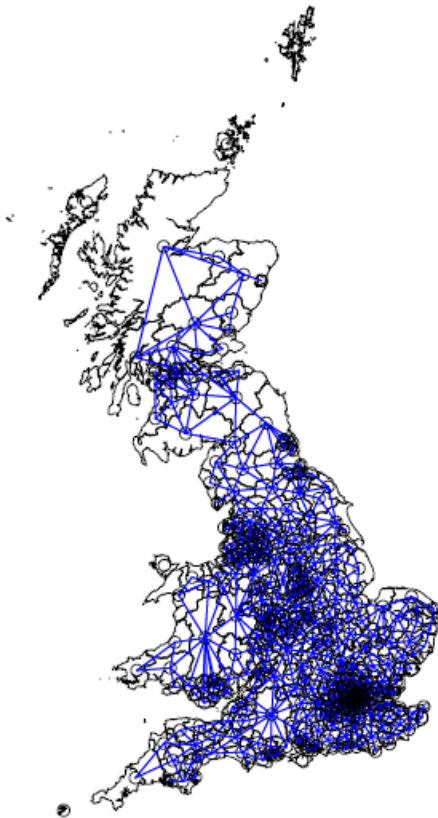
Non-competitive mimicry

Administrative professionals learn from each other by mimicking successful peers. There are several types of relevant others. The first one is the chief executive down the road. If the chief executive from a nearby council invites a peer review our chief executive is also likely to invite one. I consider this to be *non-competitive mimicry*. There is not a competitive rationale behind this kind of spill-over. Council A and council B do not compete for resources or investors with each other. Chief executive A copies the behavior of chief executive B because this is a straight forward thumb rule, or ecologically rational heuristic (Gigerenzer and Goldstein 1996) for simplified decision making. Such spatial spillover effects have been reported for policy adoption in the US (Berry and Berry 1990), and China (Zhu and Zhang 2015), among others.

I posit a simplified hypothesis: Council A is likely to invite a Peer Review if a nearby council did so. Every council is surrounded by a varying number of other councils. Figure 2 displays a spatial neighbor map for all local authorities in the UK to demonstrate the logic; from the geographical center of each council a blue line is drawn to every spatial neighbor. This results in a complex net of spatial relations (Figure 2). I argue that if a large share of surrounding councils has already invited a Peer Review, the given council is likely to invite one as well. The underlying causal effect is mimetic isomorphism.

> *Hypothesis 3* (Simple peer effect / mimetic isomorphism): The number of invited Peer Reviews in surrounding council will be positively affected with the likelihood of inviting a peer review.

Figure 2: Map of Spatial neighbors

Notes: Own figure.

The Corporate Peer Challenge Program

I will now briefly describe the English context of my further analysis. In 2010 the English Local Government Association, LGA, launched the so called Corporate Peer Challenge Program (CPC). Between 2010 and 2015 every council had been offered a cost-free peer-based performance evaluation. The offer was part of LGA's efforts to lobby for a sector-led, that is, self-organized approach to audit and inspection of local councils. During May 2010 and November 2015 120 of these 315 local councils experienced a Peer Review as part of the Corporate Peer Challenge Program of the LGA. The decision to invite a Peer Review is made by the Head of Paid Service which is an appointed manager who leads and oversees the administrative staff of a council. The LGA also requires the political backing of the council.
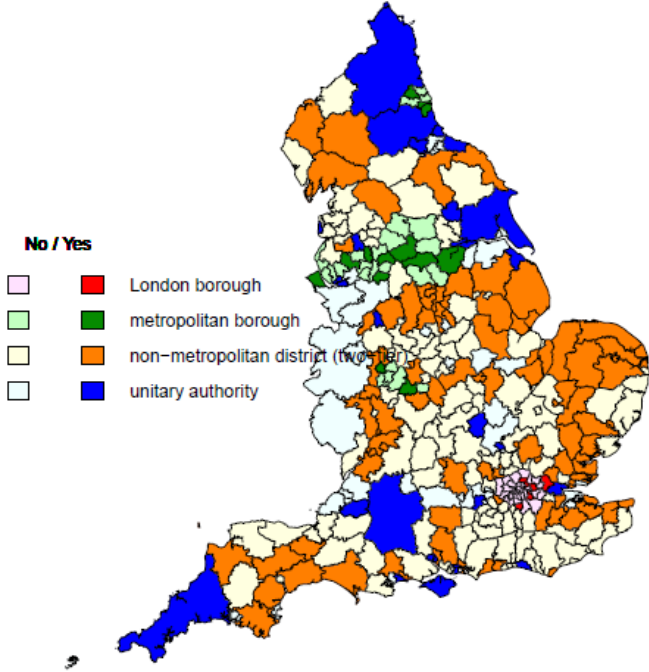
Research and data gathering underlying the paper were conducted in a two-step research process. In a first step - over the course of two months - I conducted semi-structured interviews with the chief executive, service directors and executive members from a large English metropolitan district and with LGA program directors to learn about the perceptions and views on the Corporate Peer Challenge Program.

In a second step I gathered data on the timing of the peer reviews and the composition of the review teams for 120 councils. This search took about six months. An initial list of councils that commissioned a CPC was provided by the LGA. Information on the timing of the events was extracted from individual reports that had been issued after each on-site visit. In case a council had experienced a CPC but the report or information on the timing of the event was not freely available or provided by the LGA I requested them from the council via freedom of information (FOI) requests.

Local authority districts (LAD) which commissioned a LGA Corporate Peer Challenge are spread fairly evenly throughout the country (Figure 3). Some clusters in the East of England, Southwest and the East Midlands point towards potential peer effects and/or strategic interaction in decision making. Engagement was highest among non-metropolitan districts (two-tier), and unitary authorities, and rather low among London boroughs (Figure 4, Figure 5). Timing of on-site visits: The overwhelming majority of peer reviews took place from 2012 to 2014. Demand peaked in 2013, when 54 per cent of total on-site visits were conducted

I was able to access or to successfully freedom-of-information request data for the timing of on-site visits for 120 out of 132 peer reviews. For 12 peer reviews data for the timing of on-site visits were not available for various reasons.
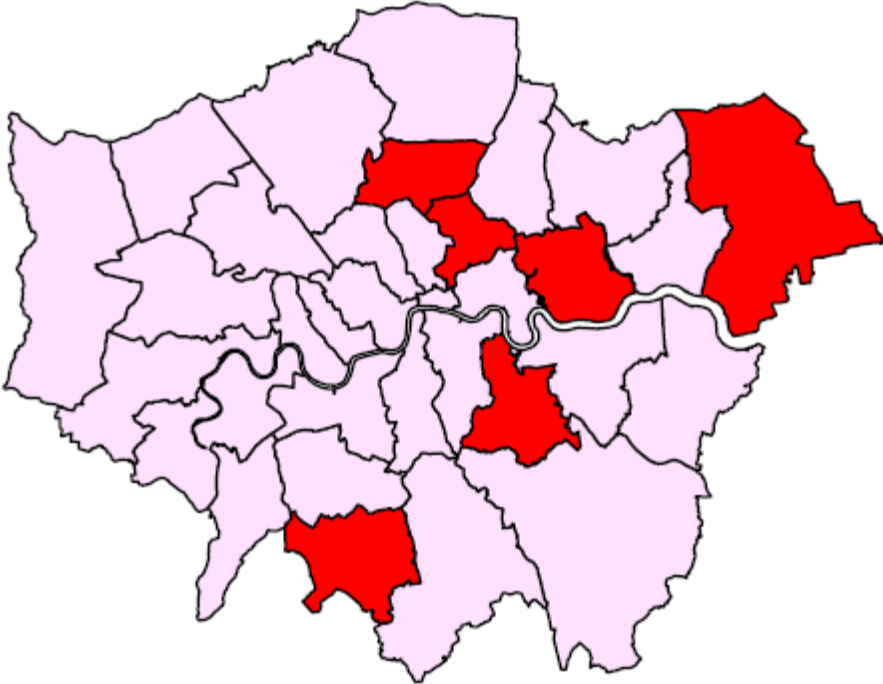
Figure 3: Map of peer reviews 2010-2015

Notes: Own figure.

Figure 4: Peer reviews by council type

Local government in England consists of four types of authorities:	The total number of each type is:	Number of LADs which between 2010 and 2015	
		invited a peer review (*as a percentage*)	did NOT invite a peer review
London borough	33	6 (18%)	27
metropolitan borough	36	12 (33%)	24
non-metropolitan district (two-tier)	201	90 (45%)	111
unitary authority	56	24 (43%)	32
Total number of LADs in England	326	132 (41%)	194

Note: Own figure and data.

Figure 5: London boroughs which issued a peer review

Notes: 6 out of 33 London boroughs invited a peer review. They are red-colored; while the administrative boundaries of boroughs without a peer review are displayed in light red. Own figure.

Figure 6: Distribution of on-site visits over time

Year	Reviews	(% total)	Councils at risk
2010	1	(<1%)	315
2011	6	(5%)	314
2012	26	(22%)	308
2013	54	(45%)	282
2014	30	(25%)	228
2015	3	(2%)	198
Total	120	(100%)	1,645

Note: Own figure and data.

Data and estimation approach

The variables of interests measure whether a council commissioned a peer review during the period May 2010 and November 2015, and if so, when the on-site review took place. The first variable is a limited dependent variable (two-choice outcome), which can take two values, 0, or 1. The sample frame consists of 315 (of 326) English local councils.

The preferred approach is an event history, or, survival regression model. There are two key concepts in event history analysis; risk set and hazard rate (Allison 2014). The risk set is "the set of individuals who are at risk of event occurrence at each point in time" (Allison 2014 p. 8). The initial risk set consists of 315 (out of 326) English local councils. The hazard rate is a "conditional probability that an event will occur at a particular time to a particular individual, given that the individual is at risk at that time" (p. 8). The hazard is an unobserved variable (p. 9).

The continuous time hazard *h(t)*, is not a probability anymore, because it has no upper bound, but it cannot be negative. *h(t)* is the unobserved rate at which an events occur. Each individual can have its one time hazard. (E.g., if 1.25 is the time hazard, this equals the expected amount of events during a one unit of time. The reciprocal $1/1.25 = .8$ equals the number of units of time until an event occurs.) Technically *h(t)* is the probability of an event divided by the time interval *s* which becomes smaller and smaller ($\lim_{s \to 0}$).

Specifying survival data

In my dataset, the variable *time* measures the working days from the start of the CPC program (May 01, 2010) and *cpc* is 1 if the council experienced an on-site visit of the review team at *time*, 0 if the council had no peer review by the end of November 2015. Saturdays and Sundays, bank holidays and all days of August (holiday season) are not counted. The values of the variable *time* range from zero to 1286 days.

Regression modeling

I specify a Cox proportional hazards model that is extended to time-variant predictors (Allison 2014). The model is parsimonious and allows for straight-forward hypothesis tests.

Log h(t) = a_t + β_1*size* + β_2*archival performance* + β_3*nearby reviews*$_{(t-1)}$ (1)

where a_t is any function of time, e.g., a constant; *size* is a logarithmic value of the population in 2012, a time-fixed variable; and *archival performance* is a score on a five points scale (1=poor, 5=excellent), time-fixed. *Nearby reviews* is the cumulated number of previous events in neighboring councils in t-1, this is a *time-variant* covariate. β_1, β_2, and β_3 are coefficients to be estimated. The model states that the log of the hazard at time *t* depends on the size of the council, its archival performance score and the value of previous peer reviews in neighboring councils at the same time. *a(t)* may be any function of time (Alison 2014 p. 35).

Size: In this study size is measured with the logarithmic value of the population. In mid-2012 an average council had a population of 164,000 community members. The minimum value is 2,264 for the Isles of Scilly, the maximum is 1,085,417 for Birmingham, England's largest local authority (London consists of 33 boroughs, with each of them being an independent local authority).

Performance: In this study performance of councils is measured with performance scores from the former Audit Commission, an independent watchdog. The Audit Commission rated the overall performance of single tier councils, unitary and metropolitan borough councils, and London Boroughs on a four-point scale, from one to four stars, with more stars indicating better performance. The Audit Commission also rated performance for 13 items and policy areas, from corporate assessment and children and young people to value for money, using the four star scheme. And the Audit Commission rated each council's overall improvement (direction of travel), the rating score ranged from not improving adequately, the worst category, to improving strongly, the best possible direction of travel. In the latest data available 3 councils were rated one star, 21 councils were rated two stars, 43 received three stars, and overall performance of 45 councils was rated four stars. District councils, and some unitary authorities, underwent a Comprehensive Performance Assessment. In a CPA performance was rated on a five point scale, from 1=poor to 5=excellent. In the latest CPA scores, one council had poor performance; eight councils were rated as performing weakly, 60 councils had a "fair" performance, 78 councils had a "good" overall performance, and 41 were ranked as excellent. To make scores comparable, the four point scale for single tier councils was standardized to a five scale item. An example: The Audit Commission rated the overall performance of Hartlepool, a unitary authority, four stars, and the direction of travel was "improving strongly". Corporate assessment was given three stars, use of resources also three stars, children and young people and adult social care also three stars. Housing in Hartlepool was rated four stars. So Hartlepool had a really good performance according to the

Audit Commission. In contrast the overall performance of Milton Keynes, another unitary authority, was rated one star, the worst category; the outlook was a grim "not improving adequately". The performance of children and young people policies was rated one star, financial reporting was rated one star, too. Milton Keynes performed well below average, according to the Audit Commission.

Nearby reviews: The cumulated number of previous events in neighboring councils in t-1 is a *time-variant* covariate. The variable gives the number of previous events in neighboring councils at a given point in time. The mean value is 0.79, the maximum value is 7.

Peer Reviews within the CPC Program are non-repeatable events. Correcting for dependence of observations from the same subject is thus not required (Allison 2014 p. 14). All events are treated equal. For estimation I use R and the *coxph* command from the survival packages (Therneau 2016).

Estimation results

Figure 7: Estimation results

Dependent variable: working days from May 01, 2010 to on-site peer review (from 0 to 1286+ right censoring)				
Variable Description	ß, a coefficient	exp(ß), a factor change	z, a z-statistic	p, a significance level
log(population in mid-2012), time-fixed	−0.25	(0.78)	−1.48	0.14
Archival performance score on a five points scale (1=poor, 5=excellent), time−0.03		(0.98)	−0.26	0.80
Cumulated number of previous events in neighboring councils, time-variant 0.05		(1.05)	0.59	0.56
Number of observations			24,605	
Wald test			2.68	
Likelihood ratio test			2.69 on 3 df	

* p<0.05, ** p<0.01, *** p<0.001

Figure 7 displays the estimation results. Hypotheses 4a to 4d are not covered by the model.

Test of hypotheses

The general model predicts that there is a relationship between the performance of a council and the invitation of a peer evaluation. *Hypothesis 1a* states that the relationship will be positive, while *hypothesis 1b* posits are negative relationship. Cox proportional hazard (COXPH) regression results at least partially support *hypothesis 1a*. The estimated coefficient has a negative sign. This suggests that councils with poor performance are more likely to invite a peer review compared to top-performing councils. However, the coefficient does not differ from zero at a significance level below 10 per cent. COXPH regression results do not fully support *Hypothesis 1a.*
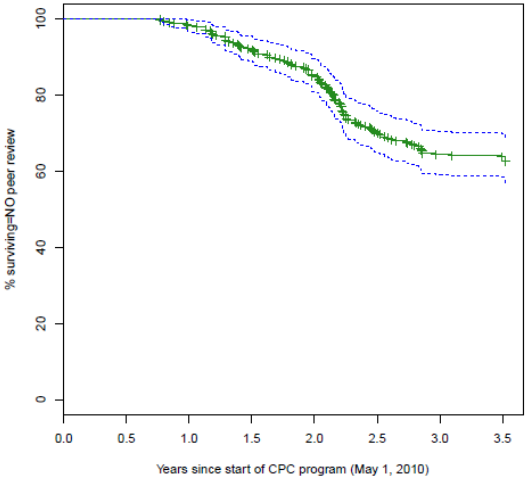
Hypothesis 2 states that size will be positively correlated with inviting a peer review. Estimation results show that logarithmic population size is negatively correlated with the hazard rate, h(t). This suggests that large-scale councils are more likely to invite a team of critical friends; the regression results thus partially support hypothesis 2. The effect is not significant at 10 per cent level, however. COXPH regression results do not fully confirm *hypothesis 2*.

Hypothesis 3 states that number of prior peer reviews in neighboring council will be positively correlated with likelihood to invite a review. COXPH regression results show that the coefficient for the variable "cumulated number of peer reviews in neighboring councils" is positively correlated with the log hazard rate. This suggests that councils which experience peer evaluations in nearby councils are more likely to commission a peer review themselves. However, the effect is not statistically significant at 10 per cent level.
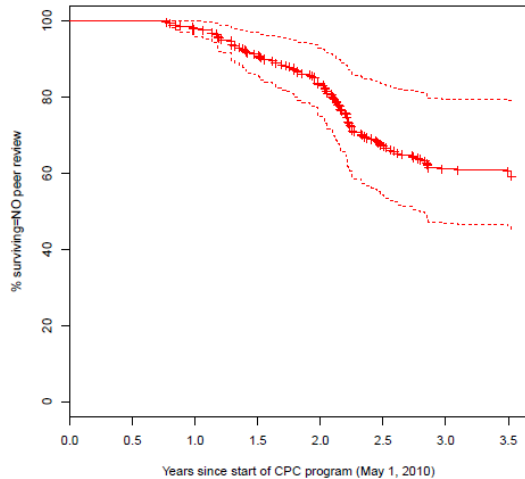
Survival distributions

Figure 8 plots the survival distribution based on the estimated cox proportional hazard model for a council of average size (some 160,000 inhabitants), one prior nearby peer review and a good performance score (pictured left). Figure 9 depicts the estimated survival distribution for the same "average" council but with a "poor" performance score. In survival analysis terms, top-performing councils 'die' later. That is, at the same point in time, a low-performing council has a higher hazard rate compared to a better performing council. This effect is not significant, however; the two distributions do not significantly differ at 10 per cent level.

Figure 8: Estimated survival probabilities: good performance

Notes: own figure. Figure 8 plots the survival distribution based on the estimated cox proportional hazard model for a council of average size (some 160,000 inhabitants), one prior nearby peer review and a good performance score

Figure 9: Estimated survival probabilities: poor performance

Notes: Own figure. Figure 9 depicts the estimated survival distribution for the same "average" council but with a "poor" performance score

Discussion

The paper addresses a crucial trade-off of administrative professionals: Acquiring new knowledge about strengths and weaknesses in organizational activities versus using existing knowledge stored in organizational memory. I argued that inviting a peer Review is a

reasonable proxy for acquiring new knowledge. I further argued both participation and non-participation in a peer review are risky choices, in terms of performance reputation. I developed a model in which observed decision results from an analysis of expected gains and losses in terms of performance reputation, peer effects among nearby organizations and organizational size. And I presented an initial parsimonious estimation model. Hypotheses derived from the model were only partially confirmed however.

Further research might investigate the role of additional factor that have not been included in the presented regression model, yet. First the model presented in this paper assumed a non-competitive environment. Future research might consider a competitive environment in which councils compete with each other for grants, investments, and resources. In such an environment a given council competes with comparable councils; it does not compete with the small village down the road, but a city of similar size and demographic structure, such competitive peers are more distant, normally. Small distance between such competitive peers intensifies the competition. Two hypotheses might be derived:

- A large number of relevant competitors having invited a peer review is positively associated with the chance that a given council invites a peer review.
- The aforementioned effect increases as the distance between two competitors decreases.

Future research might test these statements empirically. The contagion effect may also be conditional on the performance of surrounding councils professionals copy successful peers, not unsuccessful ones. Second interpersonal communication with relevant peers (Arnold 2015) is likely to act as a factor determining the chief executive's peer review decision. Third policy innovators, i.e. professionals hired from outside the organization are also more likely to implement professional innovations (Teodoro 2011). Chief executives hired from outside might be expected to invite a peer review more often compared to peers promoted within an organization. Fourth being a member in a peer review team might induce inviting a peer evaluation. Chief Executives who served their professional community are likely to invite a peer review for their own organizations. Fifth it can be argued that turnover rate at senior level is negatively correlated with the likelihood of inviting a peer review. The United Kingdom has a lively job market for chief executives. The head of paid service of big council rotate quite often. Kym Ryle had been serving as CEO at Hull City Council, and then joined Shropshire as CEO, only to take the same job in Cheshire East, three years later. In Hull he

was replaced by Nicola Yates. Nicola resigned as CEO in Hull in July 2012, but become City director in Bristol.

The paper demonstrates the application of survival analysis technique to a fundamental strategic problem in public administration; it also indicates directions for future research on the issue. The empirically results presented in this paper are based on data from England. How well does the model and the results apply to different administrative settings (external validity)? Peer review exercises in public sector organizations are a common management tool also in Northern America (Ammons and Rivenbark 2008; Charbonneau and Bellavance 2015) and Europe (Kuhlmann and Jaekel 2013, Jaekel 2016, Askim et al. 2008). For example Jostein Askim (Askim 2007, Askim 2009) examined the impact of education and experience on the extent to which Norwegian councilors search for performance information when faced with decision dilemmas. This provides insights how individual characteristics affect the use of performance information. His empirical analysis reveals that, in comparison to 'backbenchers' in the council 'frontbenchers' are more inclined to search for performance information. D. Ammons and W. Rivenbark (Ammons and Rivenbark 2008) studied the use of performance information in the US-North Carolina Benchmarking project, which was established in 2005. They found that different forms of performance measures are widely used among the 15 cities. But not every city also uses this performance information to improve its own performance. The evidence from other countries suggests that the English setting has external validity.

List of references

Allison, Paul David. 2014. Event History and Survivial Analysis. 2nd ed, Quantitative Applications in the Social Sciences. Thousand Oaks: SAGE.

Ammons, David N., and Dale J. Roenigk. 2015. Benchmarking and Interorganizational Learning in Local Government. Journal of Public Administration Research and Theory 25 (1):309-335.

Ammons, David N., and William C. Rivenbark. 2008. Factors Influencing the Use of Performance Data to Improve Municipal Services: Evidence from the North Carolina Benchmarking Project. Public Administration Review 68 (2):304-318.

Arnold, Gwen. 2014. Policy learning and science policy innovation adoption by street-level bureaucrats. Journal of Public Policy 34 (03):389-414.

Askim, Jostein. 2007. How do politicians use performance information? An analysis of the Norwegian local government experience. International Review of Administrative Sciences 73 (3):453-472.

Askim, Jostein, Age Johnsen, and Knut-Andreas Christophersen. 2008. Factors behind Organizational Learning from Benchmarking: Experiences from Norwegian Municipal Benchmarking Networks. Journal of Public Administration Research and Theory 18 (2):297-320.

Askim, Jostein. 2009. The Demand Side of Performance Measurement: Explaining Councillors' Utilization of Performance Information in Policymaking. International Public Management Journal 12 (1):24-47.

Berry, Frances Stokes, and William D. Berry. 1990. State Lottery Adoptions as Policy Innovations: An Event History Analysis. American Political Science Review 84 (02):395-415.

Besley, Timothy, and Anne Case. 1995. Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition. American Economic Review 85 (1):25-45.

Charbonneau, Étienne, and François Bellavance. 2015. Performance management in a benchmarking regime: Quebec's municipal management indicators. Canadian Public Administration 58 (1):110-137.

Choi, Taehyon, and Susan Meyers Chandler. 2015. Exploration, Exploitation, and Public Sector Innovation: An Organizational Learning Perspective for the Public Sector. Human Service Organizations Management, Leadership & Governance 39 (2):139-151.

DiMaggio, Paul J., and Walter W. Powell. 1983. The Iron Cage Revisited: Institutional

Isomorphism and Collective Rationality in Organizational Fields. American Sociological Review 48 (2):147-160.

Edwards, Ward. 1954. The theory of decision making. Psychological Bulletin 51 (4):380-417.

Fernandez, Sergio, and Tima Moldogaziev. 2013. Using Employee Empowerment to Encourage Innovative Behavior in the Public Sector. Journal of Public Administration Research and Theory 23 (1):155-187.

Gigerenzer, Gerd, and Daniel G. Goldstein. 1996. Reasoning the fast and frugal way: Models of bounded rationality. Psychological Review 103 (4):650-669.

Huang, Kun. 2014. Knowledge Sharing in a Third-Party-Governed Health and Human Services Network. Public Administration Review 74 (5):587-598.

Jaekel, Tim. 2016. Benchmarking and Comparisons of Performance of Local Governments in Germany–Approaches, Advantages and Side Effects. In COMPARATIVE STUDIES ON VERTICAL ADMINISTRATIVE REFORMS IN CHINA AND GERMANY, edited by Y. Wang and G. Faerber. Speyer: German Research Institute for Public Administration Speyer.

James, G. March. 1991. Exploration and Exploitation in Organizational Learning. Organization Science 2 (1):71-87.

Kramer, Robert. 2015, manuscript currently under review. From Skillset to Mindset: A New Paradigm for Leader Development of the Senior Civil Service.

Kuhlmann, Sabine, and Tim Jäkel. 2013. Competing, Collaborating or Controlling? Benchmarking-Regimes in European Local Governments from a Comparative Perspective. Public Money & Management 33 (4):269-276.

March, James G., and Herbert A. Simon. 1993. Organizations. 2nd ed. Oxford: Blackwell.

Simon, Herbert A. 1997. Administrative Behavior: A Study of Decision Making Processes in Administrative Organizations. 4th ed. New York: The Free Press.

Teodoro, Manuel P. 2011. Bureaucratic Ambition: Careers, Motives, and the Innovative Administrator. Edited by K. J. Meier and L. J. O'Toole, Johns Hopkins Studies in Governance and Public Management. Baltimore: The Johns Hopkins University Press.

Therneau, Terry M. 2016. Package 'survival'. Retrieved from https://cran.r-project.org/web/packages/survival/survival.pdf, Date of access 2017-02-08.

Wilson, James Q. 1989. Bureaucracy: What Government Agencies Do and Why they Do It. New York: Basic Books.

Zhu, Xufeng, and Youlang Zhang. 2015. Political Mobility and Dynamic Diffusion of

Innovation: The Spread of Municipal Pro-Business Administrative Reform in China. Journal of Public Administration Research and Theory 26 (3):535-551.

Contact details and disclaimer:

Tim Jaekel
National Research University Higher School of Economics: School of Public Administration; Assistant Professor. 20 Myasnitskaya Ulitsa, Moscow, Russian Federation, 101000.
E-mail: tjekel@hse.ru

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.