



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Igor Sheiman

INTEGRATED HEALTH CARE PAYMENT METHODS: TYPOLOGY, EVIDENCE AND PRE-CONDITIONS OF IMPLEMENTATION

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PUBLIC ADMINISTRATION
WP BRP 18/PA/2014

This Working Paper is an output of a research project implemented at the National Research University Higher School of Economics (HSE). Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

INTEGRATED HEALTH CARE PAYMENT METHODS: TYPOLOGY, EVIDENCE AND PRE-CONDITIONS OF IMPLEMENTATION

Many countries have recently started the search for new payments methods with the specific objective to encourage integration in health care delivery – teamwork of providers, their coordination and continuity of care. This paper suggests the typology of three major integrated payment methods – pay-for-performance, episode based bundled payment and global payment. A brief overview of these methods in the USA and Europe, including Russia, indicates that there is still no strong evidence of their effects on integration and other dimensions of medical service delivery performance. It is argued that relative to other integrated methods global payment is the most promising method, since it provides incentives for comprehensive organizational changes. The major pre-conditions for global payment implementation are risk bearing in integrated networks, shared savings schemes, performance transparency system, infrastructure for coordination and collaboration. It is also argued that global payment is hard to implement – mostly due to a high probability of excessive financial risks placed on providers in integrated networks. The activities to mitigate these risks are discussed based on the approaches piloted in Russia.

Key words: medical service integration, integrated payment methods, pay-for-performance, episode based bundled payment, global payment, fundholding.

JEL Classification: Z.

¹ Professor at the National Research University High School of Economics. Moscow, Russia.
igor.sheim@g23.relcom.ru

Introduction

A key instrument to encourage integration of medical service delivery is the adoption of the payment system that incorporates financial incentives for better interaction between health care providers. Many countries have recently started the search for new payments methods with the specific objective to provide such incentives. The scope of activities ranges from small-scale pilots to regional schemes.

The search for the new methods seems to be most intensive in the countries with the dominance of fee-for-service payment (FFS) and the lack of regulatory mechanisms of integration. The most interesting and ambitious projects have started in the USA and then followed by some European countries, including post-Soviet countries in transition.

The starting point of new methods discussion is a widely accepted view that FFS leads to fragmentation of care and inefficient resource allocation. For example, Brantes et al (2009) argue that FFS discourages collaboration of providers and active efforts to reduce avoidable complications of care. Halvorson (2009) presents impressive examples of disintegrating and costly provision of care due to “selling units of care” rather than “a total package of care” that can be designed to avoid aggravations of chronic disease and hospital admissions. Comparisons of physicians paid under FFS and capitation demonstrate that rates of elective surgery, patients consultations, diagnostic services, specialist and hospital referrals are substantially lower under capitation. (Dusheiko et al 2006).

The policies for health care integration have prompted a number of questions. Which methods of payment can encourage integration? What is the typology of these methods? What is the evidence of their implementation? What are the implementation problems?

This paper attempts to answer these questions. First, definition and typology of integrated payment methods are discussed. Second, a brief review of the evidence of these methods is presented. It is based on the projects implemented in the USA, some European countries and Russia. Third, relative strengths and weaknesses are evaluated against five criteria. Finally, we discuss the pre-conditions of global payment implementation as the most comprehensive economic approach to strengthen integration.

Definition and typology of integrated payment methods

What specific characteristics of integration should be promoted through new methods of payment? We suggest three major attributes: 1) teamwork of various providers, 2) coordination of their interaction, 3) continuity of care at various stages of service delivery. They most comprehensively reflect integration activities: multispecialty groups of providers are established that work on the basis of joint clinical guidelines; their members interact with each other to

enhance clinical and economic outcome; every stage of patients' "route" in the health system is highly connected with the previous and subsequent ones (Sheiman, Shevski, 2014).

The methods that are specifically designed to reach these attributes of integration can be regarded as integrated payment methods. The major expectation is that they incentivize providers to focus on integrated activities rather than selling units of care.

There are attempts to classify integrated payment methods. In the recent review of "Eurohealth" (Hernández-Quevedo et al, 2013) financial incentives and payment models are distinguished. Financial incentives are based on pay-for-performance method (P4P). The bonuses are paid for specific integration activities, mostly for achieving outcomes under chronic disease management programs. Payment models are addressed as bundled payment that is defined as "an annual payment for the complete package of care required by patients with chronic diseases".

Mechanic and Altman (2009) are more specific about bundled payment in the context of the USA health system. They suggest two types of such payment methods: bundled payment for episode of care and global payment. Bundled payment in the narrow sense is related to the specific disease or episode of care, such as hospital admission. Under this financial scheme, the rate of payment covers the cost of a bundle of services, including readmission. Global payment is defined as "all-inclusive payment per enrollee for a defined scope of services, regardless of how much care is actually provided". The authors refer to capitation method that is widely used in Europe for paying primary health care (PHC) providers and extend this method to integrated networks of outpatient and inpatient care providers.

The latter classification seems to be more precise because it takes into account the differences between methods of bundling the unit of payment. We see three major distinctions between global and episode based bundled payment. First, global payment is related to general medical activities, while bundled payment is the reimbursement of disease-specific activities. The further examples of global payment in the USA and Russia are comprehensive schemes that cover the cost of all types of care provided by integrated networks. Second, global payment implies enrollment of patients with a specific integration scheme, while bundled payment does not. It can be used together with a traditional FFS method, which is the case in widely cited Dutch chronic disease management programs (Struijs and Baan,2011). Third, global payment is always made for a specific period of time, while bundled payment is usually related to an episode of care that does not have clear cut time dimension.

Thus three integrated payment methods can be discussed – P4P, episode based bundled payment and global payment. Each of them has its own subcategories.

Pay-for-performance (P4P) can be regarded as integrated payment method with some reservation. It is not a distinct method and is always used together with other payment methods (FFS, capitation or bundled payment) to reward specific integration activities. Bearing in mind a growing interest in P4P worldwide, it is considered in this paper as the integration tool used separately or to complement other integrated methods.

Bonuses are paid to promote team care arrangements, the use of new IT for better information exchange between providers and providers and patients, recruiting and enrolling patients into chronic disease management programs, providing continuity of care. Usually financial incentives are used to influence the structure and processes of integrated care. But there are examples of their focus on the outcome of care (see the following section).

P4P can be also divided into general and disease-specific. The former rewards general integration activities, the latter – reaching targets related to specific disease-based management programs.

Episode based bundled payment. The specification of the bundle of services differs in the literature. Struijs and Baan (2011) address bundled payment as the payment for integrated set of services provided by outpatient care teams - GPs and outpatient specialists under the programs of specific chronic disease management. Brantes et al (2009) see this method as a reimbursement of inpatient care cost - not only the episode of hospital admission but also “a set period of management of a chronic condition”, including readmission caused by low quality of inpatient care. Thus this method is used for reimbursement of inpatient care or both outpatient and inpatient care.

A broader definition is used for the version of diagnostically related groups - Medicare Severity Diagnostic Related Groups in the USA. This is a flat payment weight for multi morbidity cases – for principal diagnosis and up to eight secondary diagnoses. Allowing for severity of illness provides an incentive for hospitals to improve clinical integration of various specialties (Durand-Zaleski and Obrecht, 2008).

The US Agency for Healthcare Research and Quality report suggests even broader definition of bundled payment as “a method in which providers are related to the predetermined expected costs of a grouping, or “bundle” of related health care services”, that is includes all types of bundling (Agency for Healthcare Research and Quality, 2012).

Thus all definitions agree on the object of bundled payment – episode of care with a specific diagnosis. But they differ in terms of the scope of services covered (inpatient and outpatient care or only outpatient care), coverage of diagnoses (single case vs. multi-morbidity

case), the time period for payment under a single rate (the episode of admission or a lengthy period of treatment after admission, for example, readmission).

Global payment assumes financial accountability of providers working in integrated networks. They are responsible for deviations of actual and expected cost (for example, the cost under capitation scheme). Global payment is always designed in the way that providers can keep savings and therefore are incentivized to more integration and more control over overutilization of services. They may bear risks of overspending as well.

Global payment schemes differ according to the level of financial risk bearing and the actors that act as risk bearers – PHC providers, hospitals or the entire network of providers. The specific type of global payment is a *PHC provider-fundholder scheme* (further - fundholding). Under this scheme PCH providers become holders of funding for outpatient and inpatient care. They are paid by all-inclusive capitation method, then act as purchasers of care – commission and pay to specialists, hospitals and other providers of care that deal with the enrolled population. Their risk bearing creates incentives for closer links with other providers to avoid overutilization of costly services. This method originated in Russia in late 80-s and then used or piloted in various versions in the UK, Italy, Sweden, Finland, Estonia (McCallum et al, 2009).

P4P, episode based bundled payment and global payment can be used in combination as mixed method. The options of combination can be different.

Summing up the above, we suggest the following typology of integrated payment methods:

- 1) Pay-for performance that reward:
 - general integration activities (teamwork, coordination, etc.)
 - disease specific integration activities (structure, process and outcome)
- 2) Episode-based bundled payment for:
 - outpatient care
 - outpatient and inpatient care
 - readmission
 - multi-morbidity case
- 3) Global payment with major risk bearers:
 - PHC providers (fundholding)
 - hospitals
 - all providers in integrated networks
- 4) Mixed methods
 - episode-based bundled payment + P4P

-global payment + P4P

-global payment + episode-based bundled payment for specific diseases +P4P

Evidence of integrated payment methods

Integrated methods are usually designed to achieve many objectives, including strengthening integration of medical service delivery. With the available instruments of evaluation it is hard to disentangle the impact of these methods on integration. Therefore, the following review is focused on three major dimensions of providers' performance: a) quality of care, b) utilization of care and health spending, c) integration activities.

Pay-for-performance

The most famous example of P4P in Europe is the Quality and Outcomes Framework (QOF) in **England**. The scheme was introduced to improve the quality of primary care and to encourage improvement in chronic disease management. A set of indicators and targets are used to measure GPs performance and financial rewards are linked to the actual achievement of each target.

The evidence of eight years of implementation is still inconclusive. Steel and Willems (2010) in their review of 35 studies conclude that quality has improved each year approximately in line with pre-existing trends. But there have been more significant improvements for some conditions such as diabetes and asthma. Also, there is some evidence that QOF has contributed to better monitoring of patients with chronic conditions through improvement in GP practice procedures – such as data recording, electronic records, effective alerts or reminders.

However, there are no signs that QOF contributes to integration through promoting continuity and co-ordination of care (Ham *et al* 2011). There has been some reported concern that the QOF can potentially result in the neglect of non-incentivized areas (Marengo, 2013). Also, it does not provide incentives to develop new ways of delivering care for people with co-morbidities and long-term conditions. There is a risk that people with multi-morbidity are not treated holistically, since indicators and targets are used separately for each condition, and no single organization takes the responsibility for the total of patient's needs. We agree with the concern of Appleby *et al* (2012): “ while incentives may integrate along care pathways they create a different form of fragmentation for the patient”.

Similar arrangements to promote better monitoring chronic patients have been piloted in some regions of **Russia**. But contrary to most European countries, more general indicators and targets are used for PHC providers – the coverage of check-ups, hospital admission and emergency care rates for chronic patients, home death rate, etc. These indicators are designed to

track comprehensive performance of PHC providers rather than disease specific clinical activities. For example, high rate of emergency care calls by chronic patients is considered as the “defect” of district physicians and general practitioners’ work, and therefore is highly weighted in a formal “model of performance assessment” (Sheiman, 2010).

There is a patchy evidence of the progress in these indicators, although the impact of P4P remains unclear. There are signs of better monitoring patients at risk to avoid complications of the case. The use of broader indicators creates incentives for improvement of clinical practice and provides more holistic approach to managing patients.

However, the scale of these projects in Russia is limited and they are not supported by sufficient funding. They don’t address other characteristics of integration, like teamwork, coordination, continuity of care. Also, methodological problems remain unresolved, of which the most important are the following: a) how to set targets – based on the average or the best practice, b) how to adjust for risks variation in catchment areas, c) what should be rewarded – reaching the targets or approaching them. The major implementation problem – how to ensure sustainability of P4P. New incentives tend to go out in the situation of decrease in funding.

In some countries the objective of organizational integration is articulated explicitly in P4P schemes. In *Australia* targeted providers, especially GPs, are rewarded if they form Team Care Arrangements (TCAs) for the treatment of chronic diseases. TCAs are defined as collaborative plans that involve the GP and at least two other health or care professionals, such as practice nurse. The bonus is paid for performing activities related to a patient’s health plan. There is the evidence of expanding the scheme, as well as improved patient satisfaction. The scheme is criticized for the focus on the process of planning integrated care rather than implementing a real behavior change by providers (Young, 2013). Similar arrangement is underway in *Ontario, Canada*. Physicians are rewarded to work in Family Health Teams which enable GPs to work in cooperation with specialists to treat chronically ill. Bonuses are paid for reaching preventive care targets – in addition to the revenue collected from FFS (Bienkowska-Gibbs, 2013). In *Denmark* telephone and email patients’ consultations are encouraged by bonuses to promote on-going tracking chronic patients health status, proactive managing them, coordinating care with other providers (Hernandez-Quevedo, 2013).

Although comprehensive evidence of P4P impact on integration is limited, we suggest that rewarding the achievement of specific targets indicators of integration, although useful, can hardly contribute substantially to integration. The special scheme for teamwork may promote establishing formal teams of providers, but it is not enough to ensure collaborative models of provision. In Russia and some other post-Soviet countries, the major organizational form of PHC provision is a multispecialty polyclinic which presumes teamwork. However, the evidence of

Russia indicates that formal arrangements can't guarantee collaborative efforts: GPs in polyclinics very rarely develop plans of chronic patients management jointly with specialists, the information exchange is limited, physicians don't know much about their chronic cases, etc. (Sheiman and Shevski, 2014). To make team work effective, new organizational standards are needed. They should be encouraged by more comprehensive incentives rather than P4P.

Similarly, rewarding the use of electronic recording and information exchange is not enough for the efficient use of IT. Patients data base seems to be the necessary but not the adequate condition for the integration. The regulatory requirements are needed how to use this data for strengthening contacts between providers, as well as for ensuring continuity of care. Information on the current health status and utilization of medical services must reach PHC providers together with clear rules and algorithms of activities how to respond to the data. The new data should be a signal for specific activities, for example, closer contacts between GPs and hospital doctors to deal with patients discharged from hospitals; interactions between hospitals and rehabilitative units; undertaking proactive contacts with patients.

Episode-based bundled payment

Desirable outcomes of this method are twofold. The first is to encourage a shift of care from inpatient to outpatient units of hospitals. Having a bundled rate that covers cost of admissions, physician and other clinical services, hospitals are interested to enhance the readiness of patients for admission – to have all necessary consultations and tests, plan clinical activity prior to admission. At the same time, they are interested to reduce unnecessary physician and ancillary services. The second outcome is to reduce complications and readmissions. To achieve this, the rate of payment includes expected costs for readmissions within a specified period of time. This rate is set on the basis of targeted reduction in readmissions so that hospitals will not have an “umbrella” for the reimbursement of low quality care.

The most widely cited example of bundled payment is a Geisinger Health System (GHS) initiative in the *USA*, implemented by an integrated delivery network of physicians and hospitals. This system offers forty specific clinical processes related to managing patients after coronary artery bypass surgery. The integrated rate includes the cost of surgery, all necessary tests and post discharge follow-up of patients within 90 days. The cost of services related to possible complications and readmissions is covered by this rate and not reimbursed additionally to the regular admission. The rate assumes that GHS will reduce the historic frequency of complications by half, that is pricing is based on growing requirements to quality of care (Mechanic and Altman, 2009).

Evaluation of the first eighteen months of the project implementation found a 44 percent readmissions reduction, shorter average length of stay, and reduced hospital charges. According to Geisinger executives, the success of the initiative is due in large part to integrated delivery network with an electronic record that allows to track patients after discharge. The major problem is how to equitably distribute episode payments across physicians, hospitals and other providers involved in the project. Therefore, there is some resistance of physician-hospital integrated networks to bundled payment (Mechanic and Altman, 2009).

The example of bundled payment for multi-morbidity cases is the Prometheus Payment Model in the USA. It bundles services and provides a budget with three components: base payment, patient-specific severity adjustments and allowance for potentially avoidable complications (PACs). The attempt was made to separate the costs attributable to patient-related factors from those attributable to providers' actions, for example, the cost of admission of a patient with uncontrolled diabetes or the readmission for a wound infection of a patient who has recently been discharged after cardiac bypass surgery. PACs are estimated as 22% of all private-sector health expenditure in the USA. To make these savings actual, the model provides bonuses to the hospital and physicians for working together to avoid complications and readmissions (Brantes et al, 2009).

In Europe episode based bundled payments are piloted in various countries, including the Netherlands, the UK, Denmark. Contrary to the US initiatives, they are implemented mostly in the sector of outpatient care. For example, in *the Netherlands* the bundled payment was established in 2007 for diabetes care. Under the scheme, health insurers are able to purchase a 'bundle' of services needed to manage chronic diseases through the payment of a single fee to newly created contracting entities called "care groups" These groups are clinically and financially responsible for all patients enrolled in the diabetes care program. The main objective of the bundled payment model is to encourage providers to improve coordination of care and to decrease utilization of expensive specialist care.

Evaluation of the scheme made in 2012 indicated that 25% fewer bundled payment patients utilized specialist care in comparison to care-as-usual patients (not enrolled with care groups). This contributed to some savings per patient in the cost of diabetes-specific specialist care, but when non-diabetes costs are included total specialist costs for bundled payment patients increased more than the costs for care-as-usual patients. The higher costs for bundled payment patients can be attributed to start-up costs of these groups. Also, the long-term effect is possible, since diabetes complications often take a long time to develop. This it is too early to draw definitive conclusions about the long-term impact on the cost and quality of care.

Another result of this evaluation is the revealed problem of uncertainty of cost allocation across programs of patients management and regular set of services reimbursed through FFS. There are concerns that care groups may be double billing the insurer for the same services – through traditional FFS and bundled payment. When more chronic diseases are added to the bundled payment schemes it will be difficult to determine under which bundle certain services should be billed (Liano, 2013).

This implementation problem, in our view, highlights a broader challenge – the narrow scope of bundled payment, its limited impact on integration. The scheme is designed for the specific diseases management programs and doesn't create incentives for providers in other clinical areas. There is a potential for providers to skew their activity to the most “rewardable” schemes.

Moreover, integration efforts in one clinical area promoted by bundled payment may co-exist with fragmentation of care in other clinical area. To cope with this, more holistic approaches to payment are needed, such as global payment that is related to the entire medical activities of integrated networks.

Global payment

The example of global payment is the Alternative Quality Contract (AQC) in Massachusetts, *USA*. Starting from 2009, this contract is negotiated between the insurance company and groups of providers. A 5-year global budget is provided to these groups. It covers the entire continuum of care, including inpatient, outpatient, rehabilitation and long-term care and prescription drugs. Providers in a global budget model not only share in savings if spending is below the budget but are also accountable for deficits if spending exceeds the budget. Thus they become risk bearers. The insurance plan requires enrollees to designate a primary care physician and assumes referrals for care by a specialist. The PCP's organization is accountable for all enrollee services, regardless of whether the enrollee receives care from this organization or any other provider. As of 2011, a total of 12 groups assume risks under the AQC system. Groups range from large physician-hospital organizations to small independent practices united by common leadership (Chernew et al, 2011).

The major expectation of the model is that it provides incentives against overutilization, the major concern is that it may encourage physicians to withhold care. To cope with the latter risk, the AQC groups are eligible for P4P bonuses up to 10% of their budget with performance measures of outpatient and inpatient care. Thus the entire arrangement is a combination of global payment and P4P.

Song et al (2011) compares the AQC system with control group for the period 2006-2009. The results of the evaluation demonstrate small savings (1,9% per quarter). They were achieved through changes in referral patterns to providers with lower fees rather than decrease in utilization. The AQC system is associated with the improvement in measures of quality of management of chronic conditions in adults and pediatric care, but not for adult preventive care. All AQC groups met 2009 budget targets and earned surpluses. The authors conclude that the AQC system demonstrates moderate positive results although not so substantial as was expected. The long-term impact on spending depends on future budget targets and providers' ability to further improve efficiencies in practice. The impact of the scheme on the level of integration has not been evaluated.

In *Russia* the version of global payment known as “polyclinic as fundholder” (further – fundholding) is implemented in 10 out of 83 regions of the country². Polyclinics - big multispecialty outpatient care settings - are capitated for the catchment area population, and pays for referrals to hospitals and other providers. Global payment to polyclinics includes the expected cost of outpatient care, all or part of inpatient care and care provided by special emergency care units (in Russia most of them are free-standing providers with a relatively high volume of curative responsibilities). The range of services differs in the regional schemes from small scale global budgets (e.g. in Kaluga region - only for outpatient care) to all-inclusive payment (in Kaliningrad region). Polyclinics can keep savings, therefore are interested to increase their own volume of care and reduce referrals to hospitals. Similarly to the AQC system in the US, polyclinics are accountable for deficits of finance, but contrary to AQC, formal networks with hospitals are not established, therefore polyclinics act as sole risk bearers.

The potential of fundholding to strengthen integration is based on the assumption that polyclinics are interested in preventive care, managing chronic cases (to avoid admissions and emergency care visits), strengthening coordination function of PHC providers, cooperating with high quality hospitals and ensuring continuity of care provided by various providers. This motivation may lead to controlling overutilization of care and enhancing quality and outcome of care.

The practical implementation of the scheme over the last 3-4 years in Russia has had mixed results. On the one hand, polyclinics as fundholders tend to provide more care to chronic cases. The first innovation is to set up physician-nurses teams for home visits in case of emergency or expected aggravation of health status. For example, in Perm region the number of home visits per capita by polyclinics' personnel has doubled over the first 9 months of the new

² A region is a big area with the population ranging from one to a few million people. Most of providers are state owned and governed by regional health authorities.

method implementation, while the number of emergency calls by chronic patients has decreased by 3 times. The second new approach is to set up “schools of diabetes” and “schools of asthma” (simplified versions of Western programs of chronic disease management programs).

In some regions with fundholding scheme polyclinics have attempted to be involved in planning and controlling the volumes of inpatient care. They draft the plans of their patients needs for hospital admissions and then discuss them with hospitals. The case of Samara city has proved that such an involvement can serve as a strong leverage to avoid inappropriate admissions, decrease length of stay in hospitals and ensure the higher readiness of PHC providers for the treatment after hospital admission. However, hospitals oppose to such mode of planning, therefore it has been implemented mostly due to the administrative pressure of the city government. Although promising, this approach has not been rolled out in other regions with fundholding (Sheiman, 2011).

There is evidence of fundholding impact on utilization of services. The review of the Russian regions with this method of payment demonstrated a decrease in hospital admission rates, length of stay and frequency of emergency care calls – as the first year reaction to fundholding introduction (Sheiman, 2011). However, fundholding has not contributed much to integration of care yet. Polyclinics don't control patient flows due to weakening the referral system and low trust of patients in district physicians (they act as gatekeepers). Interaction between outpatient and inpatient care providers remains low. Information exchange is in the initial stage. Potential strengths of this payment method are mitigated by inherent drawbacks of the organization of service delivery, particularly the low role of district physicians and poorly regulated interactions between providers of all levels. Also, there are conflicting incentives for integration on the part of polyclinics and hospitals. The former are interested in more cooperation with hospitals to avoid aggravations of chronic cases, while the latter are not. They tend to increase admissions, including those which are inappropriate (Sheiman, Shevski, 2014).

The special problem of fundholding is the incentive to underprovide care. Usually this is the result of excessive financial risks placed on PHC providers as fundholders. Discussing this problem in the context of the AQC model in Massachussets, Crosson suggests that only big multispecialty practices with more than 100 thousand enrollees can act as risk bearers. In Russia this has actually been done, but the problem of excessive financial risks still remains. Therefore, additional methods have been used to neutralize incentives for underreferring and underprovision, including (Sheiman, 2011):

- *Savings of polyclinics-fundholders can be kept by them only when performance targets are met.* Targets are designed to ‘capture’ potential opportunistic behavior of fundholders. For example, frequency of deaths at home or the number of the

reveled cancer cases at terminal stage are heavily weighted indicators in the above mentioned “model of performance assessment. Also, the aggregate indicators are used, such as specific-disease mortality, the number of new invalids, the number of approved patients’ complaints. This is a version of P4P that supplements fundholding scheme.

- *Financial penalties are used for patients that have not been timely referred to hospital.* Health insurers are responsible for revealing such cases and penalizing polyclinics. They are interested in this kind of control, since receive 10% of the financial penalty size.

- Some regions (for example, Kemerovo region) use the scheme of *risk sharing between health insurers and polyclinics*. The latter have the limit of their financial responsibility. Particularly costly cases are reimbursed directly by insurers. This scheme requires a clear specification of the scope of risk bearing by polyclinics. If they don’t control utilization of certain specialized services by patients (they tend to see specialists without referrals), then these services are not included in capitation rate and reimbursed directly by insurers.

But even these activities have not mitigated the tension in regional health systems between providers of various levels. For example, in the above mentioned Kaliningrad region the incentives of fundholding are so strong that bed capacity has decreased substantially, and a few hospitals have been closed. High requirements to appropriateness of admissions and length of stay, as well as selective referrals to the best providers, have been opposed by many hospitals and physicians. The system has survived only due to the political support of the regional Governor.

Thus we see promising trends initiated by fundholding in terms of reducing overutilization and enhancing outcomes of care. But incentives are not enough to compensate for the lack of organizational integration activities and the low PHC coordinating role. The method may contribute to both integration and disintegration of providers – depending on its design and implementation.

Fundholding has been also tested in a few European countries. There is some evidence of its potential to strengthen integration and thereby contribute to quality of care and reducing overutilization of services. McCallum et.al. in their review of fundholding benefits in European countries address the integration potential of this scheme – reducing isolation of small practices (alliances between GPs), strengthening GP’s coordinating function (referrals tend to be based on more information on quality of care in hospitals), joint chronic case management to avoid complications with the resulting decrease in inpatient care utilization. But the opposite

disintegrating effects (similar to the ones in Russia) have also been indicated (McCallum et al, 2006).

In more general terms, the evidence of global payment is still weak. At this stage it is clear that the method is hard to implement as it requires careful “tuning” of providers’ incentives to avoid their opportunistic behavior. However, the first experience in the USA and Russia gives ground to believe that this method may encourage more comprehensive integration activities than episode based bundled payment.

Discussion

The brief overview of newly introduced integrated payment methods indicates that there is still no strong evidence of their effects on integration activities as well as on other dimensions of service delivery system performance. The lack of evidence can be explained by the absence of evaluation instruments that allow to disentangle the influence of payment reform from wider developments in service delivery, initiated by provider-led entities (Struijs, 2013). However, the evidence available coupled with the experience of new methods implementation³ provide the ground for the comparison of their relative strengths and weaknesses, as well as making generalizations on pre-conditions of their successful implementation.

We summarize P4P, specific disease bundled payment and global payment according to five criteria. Four of them were suggested by Mechanic and Altman (2009) – potential to provide integration, reduce unnecessary utilization, encourage high quality care and operational feasibility. We suggest the fifth criteria - the degree of excessive financial risks of providers. The rationale behind the latter is that too high financial risks may potentially lead to unwillingness of providers to bear such risks or their opportunistic behavior – underreferring and undertreatment. The results of the evaluation are presented in Table 1.

³ The author has been involved in building global payment schemes in Kaluga, Kemerovo, Samara and Perm regions of Russia

Table 1. Evaluation of integrated payment methods based on key criteria

	Promoting provider integration	Controlling unnecessary utilization	Encouraging high quality care	Operational feasibility	Degree of excessive financial risks of providers
Pay-for-performance	*	-	*	***	-
Episode-based payment	**	**	**	**	*
Global Payment	***	***	**	*	***

Source: author analysis

Based on the overview, P4P method is relatively easy to implement. Providers don't bear any risks, they can only win but never lose from these scheme. The impact on integration, utilization and quality is low. Episode-based payment can promote integration in specific disease management programs with the potential of reducing unnecessary utilization and encouraging high quality care. It is relatively easy to implement. The possibility of excessive financial risks of providers exists but it is much lower than in the case of global payment – due to a relatively small scale of services to be covered by bundled payment.

Global payment method has the highest potential for integration and high performance of providers. It is more comprehensive, that is designed for the entire medical care rather than specific diseases management. Capitation rate that covers the services of big organizational entities, like physician-hospital groups, is the leverage to promote formal and virtual links between providers. As C. Ham (2012) puts it, “the more comprehensive the scope of capitated budgets, the more important this becomes”.

Global payment usually implies a substantial risk bearing by providers. For example, in some integrated networks in the US hospitals agree to pay out physicians from pre-capitized accounts after procedure before receiving reimbursement from payer. Thus risks to hospitals may be excessive (Rice, 2012). In Russia polyclinics as fundholders are supposed to pay for hospital care even in the situation when the revenue from the payer is not enough to cover all cost or when capitation rate is not risk-adjusted enough. Such risk bearing creates strong incentives for savings and therefore for closer cooperation of providers to achieve this. The other side of the

medal is that risk bearing may be excessive with the resulting opportunistic behavior of providers, their unwillingness to be involved in networks and even growing tension between providers, as was the case in Kaliningrad region of Russia.

We may conclude with a careful generalization that more comprehensive methods of payment create stronger economic incentives for integration but at the same time make integrated networks more vulnerable. Disease specific bundled payment is less conflicting and easier to implement but its potential for integration is relatively lower.

Another area of discussion is the specification of pre-conditions for a positive impact of global payment. The experience of Russia and other post-Soviet countries indicates that global payment per se may be neutral to any organizational changes. Moreover, it can create obstacles to changes if it is designed totally for cost containment. In the Soviet health system all providers had global budget that was input-based (e.g. cost of labor, utilities) irrespective of the actual volumes and quality of care, therefore did not create any incentives to improve providers performance. Integration was achieved only due to command-and control methods (Davis, 1988, Davis, 2010). Bearing in mind this “extreme” case of providers’ demotivation, we suggest a set of major pre-conditions for the global payment as integration instrument:

Risk bearing in integrated networks. Provider or group of providers that act as integrators are financially accountable not only for the savings but also for the deficits of revenue to cover cost of care provision. To make this risks acceptable and to prevent opportunistic behavior of providers, a set of activities is needed, like risk sharing, careful specification of the scope of services and the list of providers at risk. The case of the Russian fundholding is an example of the attempt to neutralize excessive financial risks.

Shared savings schemes. This is critical to ensure a transparent pattern of savings distribution between providers in the networks, particularly to those that are not directly involved in risk bearing but contribute to integration and better performance of the networks. If PHC group acts as the major risk bearer, then hospitals must be encouraged to work in such networks and meet their objectives.

Performance transparency system. Each provider in a network has clear cut performance targets that comply with its general objectives. P4P is used to encourage providers to reach these targets – in addition to the incentive of shared savings. To track the performance of each provider, a monitoring system is needed, as well as expanded analytics capabilities.

Infrastructure for coordination and collaboration. Investment is needed in service delivery re-engineering, information exchange, care coordination protocols and staff support to enable provider collaboration. The major organizational changes should precede the adoption of

global budget to make new incentives work. The following pattern of integration is a growing interdependence of organizational and economic mechanisms.

These conditions of potential input of global payment are country specific. Even within one country they may differ substantially. The lack of infrastructure for coordination and collaboration may be regarded as the sign for more careful and phased out transition to integrated payment method. In countries with particularly fragmented care and the dominance of FFS as payment method, P4P and specific disease bundled payment (or their combination) may be the first step to create incentives for integration. Their major task is to foster initial integration activities in specific areas. Global payment can be used in more “mature” integrated networks – after implementing key steps to prepare payment reform.

On this way to a broader unit of payment, combination of traditional and integrated payment methods becomes inevitable. Global payment is coupled with FFS for a set of priority services, mostly preventive ones. FFS is also used for the services that are beyond control of providers as risk bearers (they are not included in capitation rate) and therefore reimbursed directly by a payer. Also, the combination of any form of bundled payment and P4P becomes critical so that to implement savings sharing schemes according to performance targets. Thus any integrated method is most likely to be transformed into the mixed payment system. Reaching a viable combination of various methods is the major problem of integrated payment implementation.

Conclusion

Three payment methods can encourage integration in service delivery: P4P, episode based bundled payment and global payment. Each of them has its own subcategories. A suggested typology is focused on the distinctions between episode based bundled payment and global payment.

A brief overview of these methods implementation in the USA and Europe indicates that there is still no strong evidence of their effects on integration activities as well as on other dimensions of service delivery system performance. But there are grounds to believe that P4P can hardly contribute to integration as a sole method but can supplement the use of other methods. Episode based bundled payment has much higher integration potential but this method is related to specific diseases management programs and episodes of care, therefore doesn't create incentives for providers in other clinical areas. Global payment is the most promising method, since it provides incentives for comprehensive organizational changes. But it is hard to

implement, creates risks of growing tension between providers unless supported by a set of countervailing activities, such as tracking performance of each provider and P4P.

The major pre-conditions for the global payment as integration instrument are risk bearing in integrated networks, shared savings schemes, performance transparency system, infrastructure for coordination and collaboration. The lack of infrastructure may be regarded as the sign for more careful and phased out transition to integrated payment method.

More comprehensive methods of payment create stronger economic incentives for integration but at the same time they are hard to implement and make integrated networks more vulnerable. There is a dilemma of strong economic incentives with serious implementation problems and low economic incentives with no or few implementation problems.

Acknowledgement

This paper is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2011-2012.

References

Agency for Healthcare Research and Quality Bundled Payment: Effects on Health Care Spending and Quality. AHRQ Publication No. 12-E007-EF, August 2012.

Appleby J., Harrison T., Hawkins L., Dixon A. (2012) Payment by Results: How can payment systems help to deliver better care? The King's Fund.

Bienkowska-Gibbs T. (2013) Integrated care Programs in Canada. *Eurohealth* Vol.19, No.2.

Brantes F., Rosenthal M., Painter M. (2009) Building a Bridge from Fragmentation to Accountability – The Prometheus Payment Model. *New England Journal of Medicine* September 10 2009: 1033-1036.

Chernew ME, Mechanic RE, Landon BE, Safran DG. Private-payer innovation in Massachusetts: the “alternative quality contract” *Health Affairs* 2011.; 30: 51-61.

Crosson F. (2009) 21st-Century Health Care – The Case for Integrated Delivery systems. *New England Journal of Medicine* October 1, 2009.

Davis C (1988). “The organization and performance of the contemporary Soviet health system.” In Lapidus G, and G Swanson, Eeds., *State and welfare, USA/USSR: contemporary policy and practice*. Berkeley: Institute of International Studies, University of California Berkeley.

Davis C (2010). Understanding the legacy: health financing systems in the USSR and central and Eastern Europe prior to Transition. In *Reforms and Challenges Implementing health*

financing reforms – Lessons from and for countries in transition . Buckingham: Open University Press, 2010.

Durand-Zaleski I., and Obrecht F. (2008) In: Nolte E., Knai C., McKee M. (Eds) *Managing Chronic Conditions: Experience in Eight Countries*. Copenhagen, European Observatory on Health Systems and Policies.

Dusheiko et. Al. (2006) The Effects of Financial Incentives on Gatekeeping Doctors: Evidence from a Natural Experiment. *Journal of Health Economics* 25, no 3(2006) 449-478/

Halvorson J. (2009) *Health Care Will Not Reform Itself*. Productivity Press.

Ham C, Smith J, Eastmure E (2011). *Commissioning Integrated Care in A Liberated NHS*. London: Nuffield Trust.

Ham C. (2013) Payment systems and incentives to support integrated care. Presentation. The King's Fund 11 April 2013.

Hernández-Quevedo C., Llano R. and Mossialos E. (2013) Paying for integrated care: An Overview. *Eurohealth* Vol.19, No.2.

Hernandez-Quevedo (2013) Enhancing Chronic Care Management in Denmark, *Eurohealth* Vol.19, No.2.

Llano R. Bundled payments for integrated care in the Netherlands, *Eurohealth* Vol.19, No.2.

Maresso A. (2013) The Quality and Outcome Framework in England, *Eurohealth* Vol.19, No.2.

McCallum B., Brommes M., Robinson R., Bergman S., Palu T. (2009) The impact of primary care purchasing in Europe: a comparative case study of primary care reforms. In: *Primary care in the driver's seat?* Edited by Saltman R., Rico A., Boerma W. WHO, European Observatory on health systems and health policy. Open University Press.

Mechanic R. and Altman S. (2009) Payment Reform Options: Episode Payment Is A Good Place to Start. *Health Affairs*, **28**, No 2, pp 262-271.

Rice James (2011). *Integrated Care Delivery Models: Insights from USA and Europe*. Seminar at the National Research University High School of Economics (Moscow), 27 June, 2011.

Sheiman I. (2010) Sistemi stimulirovania povishenia cachestva meditsinskoi pomozchi. *Upravlenie Zdravoohraneniem* №2 2010.

Sheiman I. (2011). Pervi opit realizatsii sistemi "policlnika-fondoderzchatel. *Menedzher zdravoohranenia*, 2011, № 5.

Sheiman I., Shevski V.(2014) Evaluation of health care delivery integration: The case of the Russian Federation. *Health Policy* (2014), tp://dx.doi.org/10.1016/j.healthpol.2013.12.011

Song Z, Safran D., Bruce E., Landon M., He Y, Ellis R., Mechanic R., Day M. (2011) Health Care Spending and Quality in Year 1 of the Alternative Quality Contract. *New England Journal of Medicine* 365;10 September 8, 2011.

Steel N., Willems S. (2010) Research learning from the UK Quality and Outcomes Framework: a review of existing research. *Quality of Primary Care* 2010, 18 (2):117-25.

Struijs J., Baan C. (2011) Integrating Care through Bundled Payments — Lessons from the Netherlands. *New England Journal of Medicine* 364; 11. March 17, 2011.

Struijs J., Baan C. (2011) Integrating Care through Bundled Payments — Lessons from the Netherlands. *New England Journal of Medicine* 364; 11. March 17, 2011.

Struijs J. (2013) Payment reform and integrated care: the need for evaluation. *International Journal of Integrated Care*, v. 13, 23 December <http://www.ijic.org/>

Young R. (2013) Team Care Arrangements in Australia, *Eurohealth* Vol.19, No.2.

Igor Sheiman

Professor at the National Research University High School of Economics. Moscow, Russia. igor.sheim@g23.relcom.ru

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

© Sheiman, 2014