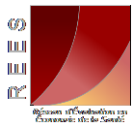
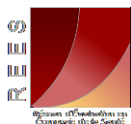


## Table of Contents

Title Page.....	3
Manuscript Statistics.....	4
Abstract.....	5
1 Introduction .....	6
2 New negotiation tools to ease the commercialization of innovative medical products .....	10
2.1 Case studies .....	10
2.2 Market Entry Agreements (MEA).....	11
3 Descriptive overview of Market Entry Agreements.....	13
3.1 Health outcome-based schemes.....	13
3.1.1 Time-bounded earmarked funding .....	13
3.1.2 Pay for performance agreements according to external standards.....	15
3.2 Financial agreements .....	19
3.2.1 Monitoring based on individual drug utilization.....	19
3.2.2 Monitoring based on population drug utilization.....	20
4 What kind of evaluation?.....	22
4.1 Descriptive studies.....	22
4.1.1 Patient registry system .....	23
4.1.2 What are registries used for?.....	23
4.2 Normative studies.....	24
4.2.1 Comparing the actual outcomes to an external standard .....	24
4.2.2 Limitations of the external criteria-reference designs .....	25
4.3 Experimental studies.....	27
4.3.1 Rubin’s canonical model .....	27
4.3.2 The strength of experimental evaluations.....	28
4.4 Quasi-experimental studies .....	29
4.4.1 Facts are not necessarily evidence .....	30



4.4.2	Analytic techniques for addressing measured and unmeasured confounding .....	30
4.5	Health outcomes vs. Performance with reference to health outcomes .....	32
5	The way forward .....	35
6	Conclusion.....	36
7	References .....	37
8	Table and figure captions.....	40
9	Tables and figures .....	41



## Title Page

### Title:

Value of Innovation in Medicine: Delimiting the Scope and Framework of Future Risk sharing Agreements

### Authors:

Robert LAUNOIS<sup>1</sup>, Lucia FIESTAS NAVARRETE<sup>1</sup>, Olivier ETHGEN<sup>2</sup>, Jean-Gabriel LE MOINE<sup>1</sup>, René GATSINGA<sup>1</sup>,

### Institutional Affiliations:

<sup>1</sup> French Network for Evaluation in Health Economics

28 rue d'Assas

75006, Paris

France

<sup>2</sup> Research Unit for the Economic Evaluation of Medical Innovation, University of Liege

Place du 20-Août, 7

4000, Liege

Belgium

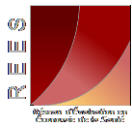
### Corresponding Author:

Name: Robert Launois

Email: launois.reesfrance@wanadoo.fr

Telephone: +33 (0)1 44 39 16 90

Address: 28 rue d'Assas, 75006, Paris, France



## Manuscript Statistics

Abstract word count: 173

Article word count: 7,027

Number of references: 46

Number of tables: 1

Number of figures: 2

## Abstract

This work presents a new taxonomy of market entry agreements (MEA), also known as risk sharing agreements. It is no longer based on the conventional distinction between outcome performance and financial contracts, proposed by Carlson. Instead, it formulates a clear distinction between monitoring studies and impact studies. These two types of study are fundamentally disparate: monitoring studies contribute to continuous program performance tracking against expected results, while impact studies seek to identify the specific effect associated with the treatment while controlling for potential sources of selection bias. As such, we seek to delimit the scope and framework of agreements that may be concluded upon between the public health authorities and the pharmaceutical industries. In accordance with this agenda, differential study designs, indicators and financial clauses are proposed to reduce clinical, economic and budgetary uncertainty. We conclude by elucidating the reasons why risk sharing contracts should be re-denominated, not as a function of virtually agreed upon treatment “performances”, but on the basis of the “effectiveness attributable to the treatment” under standard conditions of use.

# 1 Introduction

In many countries, the pharmaceutical industry does not operate within a free market. Prices are fixed through negotiations with the government and are based on clinical research findings. The main limitation of the approach is that information obtained from clinical research is ex-ante and incomplete as it does not offer insight on what happens after commercialization.

Commercializing a new pharmaceutical product presents two major risks for the payer. The first originates from the uncertainty surrounding the real value of the product. It is worthwhile mentioning that the efficacy discussed during negotiations was previously measured only in a controlled clinical environment. As such, it is not likely that the product will be as effective when targeting the general population under real world conditions. The risk/benefit ratio expected in the general population may differ from the one observed on the clinical study population. The buyer is hence at risk of paying a high price for a product that could turn out to be much less effective than advertised during negotiations [1].

The second risk arises from the unpredictability of human behavior. For a social security system to function adequately, the risk probability and the cost of its compensation should not be interlinked with individual behavior. The instability in population behavioral patterns renders risk estimates subjective and imponderable rather than predictable. Any change in behavioral patterns could jeopardize a system's ability to estimate a risk and establish counter effective measures. As such, payers are exposed to a moral hazard. The insurance companies face peril when confronted with reimbursement requests that well exceed those anticipated at the time the product was listed as reimbursable.

Various measures have been conceived in response to this problematic. Amongst them, market entry agreements have attracted considerable attention in recent years across professional and academic circles. The principle underlying such agreements is simple and works to satisfy both the industry and the payer. Before commercialization, the two parties agree on the outcome targets to be attained. If the expectations are not met, compromises are agreed upon by both parties to minimize the negative impact of unforeseeable events on either the manufacturer or the payer.

Considering the role of projections in market entry agreements, this work focuses on the processes guiding the fixation of the expected targets. There is substantial uncertainty on how to transition from clinical trial results to real world predictions and on how best to evaluate the fulfillment of those predictions. Garrison et al. have proposed the ‘performance-based risk sharing arrangements’ (PBRSA) as a solution to the issue [2]. In Europe, these agreements are referred to as ‘Market Entry Agreements’ (MEA) [3]. PBRSA (also denoted as MEA) is a practical tool enabling manufacturers to take bolder decisions with minimized risk. These contracts, which reintroduce flexibility in the drug pricing system, constitute an issue of great practical moment for the pharmaceutical industry. In this way, the regulations structuring market access procedures for pricing and reimbursement (P&R) become a unique field of research- starkly differentiated from both clinical and marketing research.

- Controlled clinical trials respond to questions relating to the efficacy of the treatment namely: “Does the therapy work?” Such trials are implemented within rigorously defined “ideal” conditions, which involve the recruitment of a homogenous population of sufferers (presenting a limited number of pre-established comorbidities and concomitant treatments) that is then followed for an arguably short period of time. Thus, the generalization of

conclusions that emerge from said trials onto the general population is perilous. This poses a problem in current clinical practice as physicians are more interested in knowing “whether the treatment is effective when administered to heterogeneous patient groups presenting the condition of interest.”

- Parallel to trials, the marketing units of pharmaceutical companies are known to produce an increasing number of descriptive studies. The main aim of these studies is to document patient complaints, as well as, the social and psychological consequences associated with the treatment. A descriptive study is characterized by the fact that it does not intervene on the natural course of treatment administration, including the dosage and duration of treatment and the terms of clinical management and surveillance. The strength of such studies is that the results emanate from real world conditions, allowing for conclusions that are more faithful to the lived experience of heterogeneous patient groups. Their major weakness is that they are not formulated with a comparator in mind.

How may we bridge the gap between experimentation and real life? The study designs described by the US General Accounting Office [4], known as normative and quasi-experimental, permit a continuum between the artificial context of experimental studies and the real-world conditions characteristic of seeding studies. Once the Market Authorization (MA) has been accorded, monitoring studies, which could either be descriptive or normative in nature, are implemented so as to verify if the results envisioned are actually obtained, while taking into account that the transposability of results obtained through controlled clinical trials is not guaranteed. Monitoring studies can expose drug consumption patterns, as well as, provide results that may guide the formulation of hypotheses to be consequently confronted through the implementation of



experimental or quasi-experimental impact studies. Impact studies could then be put into service with the aim to confirm whether the observed changes in natural disease progression are directly and exclusively attributed to the innovative treatment.

The objective of the present work is four-fold. First, we seek to delimit the scope and framework of agreements that may be concluded upon between the public health authorities and the pharmaceutical industries. Second, we will proceed to classify the relevant contractual agreements by way of their content. Third, we will elaborate on the conceptual background leading to the distinction between (i) time-bounded earmark funding devoted to the implementation of field studies and (ii) pay for performance agreements guided by pre-established indicator targets. We will conclude by elucidating the reasons why risk sharing contracts should be re-denominated, not as a function of virtually agreed upon treatment “performances”, but on the basis of the treatment’s “attributable effectiveness” under standard conditions of use.

## 2 New negotiation tools to ease the commercialization of innovative medical products

### 2.1 Case studies

The most advantageous use of a new treatment requires strict adherence to the MA label, which is neither enacted by the patients nor the prescribers. An alternative to mitigate this constitutes fixing mutual commitments agreed upon by all stakeholders, who may include: the relevant health authorities, pharmaceutical laboratories, hospitals and health practitioners. These accords maintain a contractual nature so as to guarantee the proper use of the treatment.

A good instance of this agreement is exemplified by the contract signed in 1999 between the North Staffordshire Authority and the Park Davis laboratory [2]. The manufacturer demonstrated through clinical trials that its drug, atorvastatin, produced as Tahor, could lower cholesterol levels in real life patients. They defined the efficacy criterion as the proportion of patients whose blood cholesterol levels were lowered below 3.0mmol/l. Among patients with moderate hypercholesterolemia (3.0 to 4.8mmol/l), the efficacy was 89% when administering 10mg of atorvastatin.

The manufacturer was cautious when signing the agreement, considering that the controlled conditions of the clinical trial could have led to the non-generalizability of results in an uncontrolled real life environment. The manufacturer agreed to refund the expenses associated with product use to the National Health Service (NHS) only if the efficacy observed in the trial was greater than 20% relative to the effectiveness that would be observed in real life. As such, atorvastatin would be refunded in the event that it fell under the 71% real life efficacy threshold

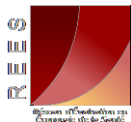
that was established a priori. An inspection funded by the manufacturer revealed the post-trial drug efficacy to be 88% and fees were not refunded.

## 2.2 Market Entry Agreements (MEA)

Prior to developing the proposed evaluation scheme in full, it is imperative to establish a common understanding of risk sharing agreements. The use of these agreements is spreading to many countries with emerging terminologies used in the literature: ‘risk sharing scheme’ in the United States [5], ‘patient access scheme’ in Great Britain [6], ‘deed of agreement’ in Australia [7], ‘market-entry agreements’ in Europe [3], and ‘access with evidence development’ in Canada [8]. This evidences the contemporary debate on a universal standard definition and categorization for risk sharing contracts. However, behind modest differences in terminology inhere fundamental disparities in the factors defining their application.

A number of academics share a restrictive definition of market-entry contracts. For instance, Carlson [9], de Pouvourville [10], and the International Society for Pharmacoeconomics and Outcome Research (ISPOR) [2] suggest that the term ‘risk sharing or market entry contract’ be reserved solely for cases where there is a plausible link between the performance of a drug and its pricing or between granted funding and collection of real world observed data. They would, otherwise, not really be ‘risk sharing’ agreements per se, but mere conventions between the industry and the national administration to share the rent of innovation. This definition systematically excludes all financial agreements.

On the other hand, authors like Adamski [11], operate with a broader view on the subject and define risk sharing contracts as: ‘agreements concluded by payers and pharmaceutical companies



to diminish the impact on the payer's budget of new and existing medicines brought about by either the uncertainty of the value of the medicine and/or the need to work within finite budgets.’

In this work, we chose to use Adamski's broader definition because it reflects the true preoccupation of the payers. The exclusion of financial contracts from the field of commitments between the industry and the payer is an arbitrary and potentially counterproductive choice. Indeed, their exclusion fails to recognize a great number of approaches that evaluate performance using the same procedures and insure the fulfillment of expected standards.

This definition brings us to distinguish two main categories of MEA agreements: health outcome-based and financially based, each of them containing respective subdivisions.

### **3 Descriptive overview of Market Entry Agreements**

The term “Market Entry Agreements” may be seen as an umbrella of health contracts that could be built upon treatment effectiveness (i.e. health outcome-based) or upon the budgetary commitments of financial contracts (i.e. financially based).

#### **3.1 Health outcome-based schemes**

The agreements included herein carry the potential of impelling pharmaceutical industries to operate in ways that satisfy the general interest. Health outcome-based schemes may be viewed as a key link between the industry and the public. A contract of this nature compels the manufacturers (i.e. “the agents” according to the principal agent theory) to provide concrete health results reflecting both their unique scientific competences, as well as, the social responsibilities shared with the payers and the rest of society [12]. Manufacturers have the scientific advantage allowing for therapeutic innovation and improvement of medical care. On the other hand, it is within the scope of the government (i.e. “the principal”) to direct research towards the public’s real and perceived needs. This is done by carefully fixing prices that incentivize industries to meet public needs. Consequently, the social and financial responsibilities of pharmaceutical companies are reconciled.

Operating within this framework, we identified two types of health outcome-based schemes in the literature: coverage with evidence development (CED) and pay for performance agreements.

##### **3.1.1 Time-bounded earmarked funding**

This term was coined by the French General Directorate for Care Provision (DGOS) to refer to the French Support Program for Expensive and Innovative Techniques (PSTIC) [13, 14]. The aims

and content of these agreements are reminiscent of the US Medicare's 'coverage with evidence development' (CED), developed in 1996 [9]. The essence of the CED agreement is that the payer temporarily subsidizes the use of the innovative treatment in order to give the manufacturer an opportunity to collect sufficient real world data. This subsidized time-window will allow the manufacturer to document the true clinical and/or medico-economic usefulness of the product under review. In lay terms, the payer agrees to a financial commitment in order to obtain additional evidence. As such, the payer may decide to limit funding so as to treat only the study participants or expand it to include all patients following treatment until the study delivers clear findings. The approach can either be experimental via the implementation of randomized trials or observational with the inclusion of a control arm<sup>1</sup> [15]. Both designs are viable as long as the study provides clear evidence that could guide decision-making on whether to continue funding the drug.

The Italian version of this scheme operates inversely as prospective observational studies are developed with the aim to verify the real world effectiveness of the innovation ex post. In these cases, the payer refrains from engaging financially until the evidence supporting effectiveness surfaces.

INSERT FIGURE 1

---

<sup>1</sup> Prospective or retrospective cohort or registry, case control studies, medico-administrative databanks, etc.

### **3.1.2 Pay for performance agreements according to external standards**

Pay for performance agreements are made when the manufacturer claims a treatment is worth more than what the payer offers. Indicators are chosen to predict the ‘performance’ of the drug on an objective and reproducible scale. As such, the obligations of each party will be defined in most of the cases according to clinical endpoints.

The payer, then, agrees to pay the price proposed by the manufacturer, on the condition that the latter would offer evidence pointing to the worth of the treatment after a reasonable period of time. If the product is revealed to be as efficient as expected, the price will be set at the amount requested by the manufacturer a priori. If this is not the case, the manufacturer reimburses the difference between the price paid and the price originally proposed by the payer.

Three types of reimbursement categories encompassed within the pay for performance scheme are presented: (1) Commitment to health outcomes in routine clinical use based on external performance standards, (2) Commitment to health outcomes in a stratified medicine context and (3) Commitment to improve medication adherence.

#### **3.1.2.1 Commitment to health outcomes in routine clinical use based on external performance standards**

Indicators are chosen to predict the ‘performance’ of the drug on an objective and reproducible scale. Negotiations between the payer and the manufacturer have the objective to establish an agreement on the target values of these parameters. They could choose to make commitments either on the final results or on intermediary results. If the decision is based on intermediary results, this “intermediate endpoint” should show a strong correlation with the clinical parameters that it

intends to replace in order to be considered a valid “surrogate” criterion. Failure of the observed indicator value to attain the target values will result in the full or partial refund of the expenditures by the pharmaceutical company.

It is important to note that the negotiations of target objectives, based on intermediary or clinical endpoints criteria, do not obey any rigorous scientific criteria. Paradoxically, these endpoints are most relevant to the population with the condition to be treated or intervened upon. In this sense, real life targets are to be established after the publication of findings from randomized trials and without any real life evidence to support the estimates. As such, hypotheses risk being unfounded, making it possible for sporadic results to be further validated by observations.

Comparing the health economic use of a new treatment with the existing therapeutic arsenal relies on the false assumption that there is a threshold for the collective willingness to pay in France [16]. Presently, this is not the case. Such threshold value could vary depending on the importance of the disease, the amplitude of unsatisfied needs, and the magnitude of the therapeutic innovation. Having threshold values would not be appalling since they are commonly used in diagnostics to distinguish diseased from disease-free populations. However, in the absence of economic threshold values, the debate guiding drug pricing would be solely grounded on intuitive medico-economic data. As such, pricing negotiations would rely on TIABIM decision-making: ‘Taking Into Account and Bearing In Mind’. As this comes to light, one can predict, that the French Economic Committee on Healthcare Products (*Comité Economique des Produits de Santé* or CEPS) will use performance-based contracts as a means to avoid cost-effectiveness analyses and to isolate the focus on financial impact analyses.



### 3.1.2.2 Commitment to health outcomes in a stratified medicine<sup>2</sup> context

Departing from the concept of a therapeutic continuum, individualized and empirical medicine are at opposite ends of a spectrum. Examples of individualized medicine are cancer vaccines based on a particular patient's tumor. On the other hand, empirical medicine represents active agents that work for most patients sharing a particular condition. The field of stratified medicine is situated at the center of this continuum. In such cases, patients are identified as part of a cohort by use of a clinical biomarker and treated with a particular therapy, for which a differential response has been established.

The advancement of clinical knowledge can help us better predict treatment response in a group of patients sharing similar demographic, prognostic and other characteristics. These could include: (a) non-responder rates defined as disease progression, progression-/toxicity-related death or unacceptable toxicity requiring treatment discontinuation in oncology; (b) glycated hemoglobin rates in diabetes studies; (c) threshold T bone density score in osteoporotic fracture prevention; and (d) Her2/neu biomarkers predicting the reaction to Trastuzumab in 20% of metastatic breast cancer patients. Knowledge of these indicators makes it possible to identify populations susceptible to respond to a treatment, while contributing to the improvement of care efficiency. Consequently, the payer is more interested in buying the treatment for patients in that category. Moreover, since the chances of an optimized response to treatment are increased, the coverage agreements are adjusted accordingly.

---

<sup>2</sup> Personalized medicine can mean various things. It can be the use of biomarkers and companion tests to tailor the best treatment for the right patient at the right time. It can also refer to the identification of sub-population groups that fit a certain profile; this is also known as 'stratified' medicine –stratified on heterogeneous treatment effects. We refer to the latter meaning when using this term.

Nevertheless, defined intermediate endpoints do not necessarily lead to clinical endpoints. For instance, in cancer studies, the correlation between the progression-free survival and the global survival is not well established. In the same way, diagnoses made using biomarkers are hindered by the scarcity of all-or-nothing predicting markers. When using biomarkers, the resulting graphical distributions of test positivity in the diseased and disease-free populations often have an overlapping area [17]. Biomarkers, much like human error, may lead to false negative and false positive results. As reported by the French parliament members, Claeys and Vialatte, the biological identification of tumors remains only partial [18]. According to the National Cancer Institute (INCa), there were only eight markers in 2012 upon which 11 targeted therapies were based [19].

### **3.1.2.3 Commitment to improve medication adherence**

Under these agreements, the coverage of a treatment claimed to improve compliance is approved with the condition that the manufacturer must monitor real life compliance and refund social security spending if adherence targets are not met. In the case referring to the commercialization of Risperidone in France, the manufacturer received its competitor's price from the beginning while the remainder premium prize was blocked into an escrow account – as prescribed within the framework of the agreement. Once real life evidence emerged [20] revealing that drug administration increased treatment response rates (i.e. improvement of medication adherence), the premium price was unblocked from the escrow account and the manufacturer received the full sum.

## 3.2 Financial agreements

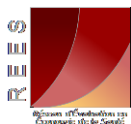
The other part of market entry agreements comprises agreements that are primarily based financial performance of new products [21]. Financial agreements aim to control the economic consequences that a negotiated price might have on the drug pricing structure, as well as, the impact that it may have on the global medical market and the national health insurance expenses [22]. These contracts are divided in two categories according to the commitments involved and are denominated as individual and target-population monitoring.

### 3.2.1 Monitoring based on individual drug utilization

As it concerns innovative practices, it is imperative to manage their spread, avoid the misuse of new technologies and limit unnecessary spending. For medicines approved for hospital use, reimbursement requires an ‘appropriate use’ contract to be signed by the Regional Health Agency (ARS), the health insurance and the hospital. To simplify the process, national French health entities (i.e. INCa, ANSM and HAS) have as mission to elaborate frames of reference for “appropriate use” within each of their domains. The notion of appropriate use covers the following four categories:

- Category I: compliance with the MA label.
- Category II: compliance with the therapeutic protocols established by any of the three aforementioned entities for indications that go beyond the MA granted
- Category III: prescriptions within the de facto and exceptional use frame, which is to be requisitioned in the presence of a strong medical argument.
- Category IV: Non acceptable situations (NAS)

Within the frame of an appropriate use agreement between the regional health agency (ARS), the public health insurance and the local hospital, the non-compliant uses of health products would

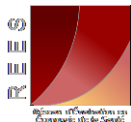


lead to the recovery of undue payment or the targeted reduction of reimbursement rates. Adherence to these rules is verified by the medical or pharmacist officer who counsels the local sickness funds, and who control approximately 5% of medical records within 10% of healthcare establishments in the region.

Monitoring patient files individually requires a complex data collection system, able to take into account the diversity of individual therapeutic paths while using numerous parameters. Monitoring is often implemented through the use of patient registries, which monitor the appropriate use of the drug for each patient. This is very time-consuming and adds tedious work onto the already heavy workload of healthcare workers.

### **3.2.2 Monitoring based on population drug utilization**

Financial contracts about populations at risk, also known as target or exposed population, are usually made to stay within budget limits. Every year, the French parliament fixes a revenue growth rate excluding taxes from reimbursable prescriptions. This is otherwise known as the « k » rate (0.4% in 2014), above which enterprises must pay a contribution called the safeguarding clause. For its part, the CEPS determines how the pharmaceutical industries may ventilate the overall rate between the various pharmacotherapeutic homogeneous classes in terms of product substitutability using the EPHMRA codes. It also fixes a rate of evolution modulated by class (given the current k rate) and dependant on the extent of unmet needs and the innovative character of the treatments that they comprise. Beyond this specific rate of evolution, clawback payments will be payable. The LEEM shall be consulted before the table is finalized. Companies may be exempted from payment of this contribution if they choose to enter an agreement with the CEPS. The safeguard clause is actually a "theoretical contribution", to the extent that almost all companies



have signed agreements with the CEPS. As part of these, they accept some price declines or undertake to pay the contractually agreed upon clawbacks. These conventional clawbacks are three: clawback payments per pharmacotherapeutic class grouping<sup>3</sup>, clawback payments based on capped turnover rates and clawback payments per product. The sum of these clawbacks may not exceed the amount which the company would have been liable for if the k rate had been applied to its revenue excluding taxes. This entire regulatory arsenal aged badly. Today, part of the device is inapplicable when the revenue growth rate is negative as is currently the case in France.

Even though the taxonomy of market entry agreements described above appears to be clear cut, most signed agreements cannot be categorized exclusively into one type. For example, the Bortezomib (Velcade) contract signed by Celgene and the NHS for multiple myeloma could be seen as a financial agreement, or a pay for performance agreement. But in fact, the contract borrows aspects from both contract types: the manufacturer commits to refunding the payer if the clinical expectations are not met (i.e. pay for performance agreement) while establishing a maximum spending limit per patient (i.e. financially based agreement).

---

<sup>3</sup> “For each of the drug groups in which the annual growth rate is higher than the rate set by the committee, the total amount owing by all the companies who have agreements with CEPS and sell drugs belonging to that group shall be equal to the difference between the two rates multiplied by a given coefficient, which shall be the same for all groups” (CEPS annual report 2010 p56)

## 4 What kind of evaluation?

When funding is allocated to the collection of new data, the goal is either to demonstrate causality between the treatment and the clinical endpoint or to examine the appropriateness of the treatment in a normative approach. In the first case, the evaluation determines the exact changes observed in patients' health outcomes that can be directly and exclusively attributed to treatment. This would constitute a scientific demonstration of causality. In the second case, whereby a pay for performance or a financially-based agreement is accorded, the evaluation becomes strictly normative. Then, the nature of the evaluation turns into an audit procedure that verifies treatment adherence against conventionally agreed upon standards.

Given the two roles that an evaluation study has, we distinguish between impact studies and monitoring studies, as suggested by the World Bank [23, 24]. Impact studies (i.e. experimental and quasi-experimental) can reveal the true added-value that an innovative health product might have over others- by aiming to establish a causal link between treatment and effect. Monitoring studies (i.e. descriptive and normative), on the other hand, are used as check-ups in order to ensure that all commitments made by the pharmaceutical industries are fulfilled, be it budgetary or clinical.

### 4.1 Descriptive studies

Descriptive studies gather and analyze data in order to verify that the available means were used as planned, that the planned activities were carried out to completion and that the agreed products and services were delivered [25]. Such studies require a set of accurate indicators that enable the evaluation of an undergoing project. These indicators should be in accordance with the SMART

criteria in order to be practical: Specific, Measurable, Attributable, Realistic and Time-bounded.

Patient registries are commonly known examples of descriptive studies.

#### **4.1.1 Patient registry system**

A patient registry is defined as “an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes”[26]. A deep understanding of the potential for bias threatening observational designs, as well as, mastery of methodological and operational instruments is required from the registry implementer. This is needed in order to minimize the influence of such biases.

#### **4.1.2 What are registries used for?**

Creating a patient registry requires: (1) defining its content: registry by product, indication, or population; (2) defining the goal of the registry: epidemiological research on the natural history of the disease, clinical research on comparative effectiveness of treatments in the real world, or utilization study tracking appropriate use of the drug for each patient; (3) verifying that information is not recorded more than once through agencies with similar missions like the Regional Health Agency (*Agence Régionale de Santé* or ARS) and the French Drugs Medical Devices and Therapeutic Innovations Observatory (OmeDIT); and (4) defining a particular term to distinguish the drug registries from other registries that focus most on epidemiology and do not aim to measure clinical performance. In France, the registries established to evaluate the natural history of the disease are the sole ones to receive government funding and are established by a commission

involving the French National Institute of Health and Medical Research (Inserm) and the French Institute for Public Health Surveillance (InVS) following the decree of May 23<sup>rd</sup> 2013.

Registries vary in accordance with their assigned objectives. When registries are used to assess the appropriateness of drug use, all drug users must be included, which simplifies data collection. When they are meant to collect safety information concerning the patient risk profile, the data collection must be bounded by a time period and a regional extension that is coherent with the incidence of the events of interest. Conversely, when they are focused on effectiveness, the main aim is to identify the prognostic factors driving differences between the pre- and post-marketing studies. In these cases, data collection may be limited to a representative subgroup of users.

## 4.2 Normative studies

### 4.2.1 Comparing the actual outcomes to an external standard

As stated in the article 1134 of the French civil code: ‘freely formulated conventions stand as law to those involved.’ If such agreements were to be signed between the French Economic Committee for Health Products (CEPS) and a member of the French Pharmacological Companies Association (LEEM), it would be as private commitments integrated as waivers in the LEEM-CEPS Framework Agreement<sup>4</sup>. These agreements are based on the use of SMART indicators, as seen with descriptive studies. However, in the case of normative studies, these indicators are enriched by a contractually-fixed target and a baseline value representing the situation before treatment

---

<sup>4</sup> Framework Agreement of the 5th of December 2012 between the French Economic Committee for Health Products and the French Association of Pharmacological Companies (*Les Entreprises du Médicament* or LEEM)



administration. If the indicators are not specified in terms of a baseline value, time frame and quantity, we cannot be sure of being on the right path towards the target set.

The benefit of having such contracts is that the expectations of each party are clearly defined [27]. The drug performance that the CEPS expects to observe in return for the agreed price and the medical and/or budgetary objectives that the pharmaceutical firm commits to achieve are fixed a priori.

One might be able to formulate an educated estimate and attempt to set up a criteria based on clinical trial results. Yet, the uncertainty of translating knowledge originating from a controlled clinical trial environment into the standard to be expected under real life conditions, could subject decision-making to pure happenstance. The term ‘medico-administrative’ used by the French Directorate General for Health Care Organization (*Direction Générale de l’Organisation des Soins* or DGOS) in a ministerial circular [28] represents a mere administrative action attempting to translate the wealth of clinical evidence into a ministerial circular.

Normative studies are limited to corroborating the fulfillment of commitments made in an agreement. They are not methodologically equipped to demonstrate that the objective has been achieved due to the treatment. Such an evaluation cannot verify that the innovation improves health outcomes in real life or increase the efficiency of the treatments.

#### **4.2.2 Limitations of the external criteria-reference designs**

A discordance is to be expected between the service that is anticipated and that which is observed. The reference is built upon controlled trials and meta-analyses, while performance is observed in and confounded by uncontrolled real life factors. The vicious circle is closed when treatments are

developed and evaluated focusing on the fulfillment of norms abstracted from controlled clinical settings. This situation is well illustrated by the zebra crossing area phenomenon, particularly in that officials are more preoccupied about its proper use rather than on how useful it actually is. This ‘administrative’ state approach leans towards the ‘normative research’ definition that A.P. Contandriolopoulos [29] and the US General Accounting office [4] refer to: ‘a judgment made based on the comparison of the used resources, services rendered or goods produced, and obtained results with certain norms and criteria.’ The fact that the norms were negotiated through a contractual agreement does not alter the purely legal and unscientific nature of their content [30]. Indeed, it is a regression to the past attempt to introduce mandatory clinical guidelines in France.

There is no certainty about the effectiveness of a treatment in real life. ‘Performance’ should under no circumstance be confounded with ‘result.’ By measuring performance, a treatment could reach the pre-established target value. However, the true effect of the treatment may not be the cause behind the value attained. This makes the performance-based evaluation of new treatments unreliable. In this vein, the evaluation of treatments becomes ‘virtual’ in every respect except for the legal aspects of the financial commitments made. To estimate the ‘true result’ that can be attributed solely to the treatment under consideration, both the analyst and the decision maker need more robust analytical tools.

The foundations of the specified approaches, namely: coverage with evidence development contracts and pay for performance agreements, are rendered clear in Carlson’s article [9] but is read ambiguously in subsequent works by Garrison [2]. Essentially, Garrison’s proposal is founded upon the overall absence of a comparator in the performance-based agreements. Though the distinction between normative evaluation and evaluative research [29] continues to exist under

different designations (i.e. real-world product evaluation versus pay-for-performance agreements “with ex ante commitments”), Garrison privileges the latter. This reasoning is based on contractual norms defined a priori that neglect the existence of potentially confounding factors, thus, guiding the prescription of individualized patient monitoring within the framework of registries. We warn against the transposition of such a proposal to the French context as it risks enlarging the State’s role in the mechanisms setting drug prices.

## 4.3 Experimental studies

### 4.3.1 Rubin’s canonical model

Evaluating a causal relationship between a treatment and a health outcome in an individual necessitates a direct comparison between the treatment-related benefit gained by the individual under treatment A and the treatment-related benefit that the same individual would have gained should she/he have taken a different treatment [20, 31]. This comparison is not a practical but a theoretical ideal because a single individual cannot simultaneously undergo two different exclusive treatments.

A counterfactual situation cannot be estimated without calculating the total treatment difference in effectiveness ( $\Delta E^{TT}$ ) between the average health outcome  $E(y_{i1})$  of patients treated with the innovative treatment ( $T = 1$ ) and the average health outcome  $E(y_{i0})$  of patients treated conventionally ( $T = 0$ ). In other words, the *effect attributable* to the treatment ( $\Delta AE^{TT}$ ) must be completely isolated from the effect of the confounding covariates ( $\Delta CE^{TT}$ ) in order to make a causal inference linking treatment and outcome directly:

$$\Delta E^{TT} = E(Y_{i1}|X_i, T_i = 1) - E(Y_{i0}|X_i, T_i = 0)$$

$$\Delta E^{TT} = E(Y_{i1}|X_i, T_i = 1) - E(Y_{i0}|X_i, T_i = 1) + E(Y_{i1}|X_i, T_i = 0) - E(Y_{i0}|X_i, T_i = 0)$$

$$\Delta E^{TT} = \Delta AE^{TT} + \Delta CE^{TT}$$

Methods have been developed in order to approach the ideal comparison standard proposed by Rubin practically. For instance, one can attempt to find a conditional ‘counterfactual’ for every treatment A recipient, by identifying a person with similar demographic and prognostic characteristics undergoing treatment B. In which case, Treatment B could either signify the absence of A (i.e. placebo agent) or a different active agent. The probability of finding an exact counterfactual is, however, extremely low. As a result, when comparing two individuals expected to have similar characteristics in every other respect (with the exception of treatment assignment), the risk of mistaking the effect of confounders for the direct effect of the treatment remains sizable. It is possible to mix two effects, one which is attributed to the treatment (i.e. the ‘appropriate effect’) and one which is attributed to factors other than the treatment (i.e. the ‘confounder effect’). The ‘confounder effect’ may be the result of preexisting health state differences between the treated and untreated subjects, also referred to as prognostic factors influencing the choice to treat. Indeed, conclusions derived from treatment comparisons plagued by known and unknown confounding factors are likely to lead to channeling bias. Further, failing to account for such factors may lead to distorting the observed treatment effect.

### 4.3.2 The strength of experimental evaluations

Double-blinded randomized trials establish a causal relation [32] via the following three conditions:

- Randomization ensures the comparability between the treated and untreated groups by neutralizing visible and non-visible personal characteristics.
- The use of placebo provides a common reference to assess the effect magnitude of alternative treatments.
- The double blind guarantees behavior comparability between treatment arms by neutralizing the effects of examiner, respondent and structural biases.

Given the aforementioned aspects of an experimental evaluation, the differences that arise between the groups could be exclusively attributed to the treatment. When taking these conditions into account, the odds ratios and risk differences obtained would no longer measure a mere association but the amount of true effect attributable to the treatment. Thus, proving a causal relation.

#### **4.4 Quasi-experimental studies**

In the absence of randomization, risks for bias rise. Bias is a systematic interference corrupting the statistical inferences that can be made from sampled data. Accordingly, the estimates of a certain parameter studied from the sample may not be representative of its real value in the population. The risk of bias persists through each step of a non-experimental study: from sampling (i.e. selection bias), to data collection (i.e. information bias), and data analysis (i.e. confounding factors<sup>5</sup>).

---

<sup>5</sup> A confounding factor is a mixture of effects (third and exposure factors) that biases the interpretation of the study results due to the simultaneous association of the third factor: (1) with the exposure (i.e. treatment); (2) and with the occurrence of the event mislead to the belief that the exposure triggered the event. Ex: Grey hair/ infarctus/age

#### **4.4.1 Facts are not necessarily evidence**

The existence and proliferation of bias is innately tied to observational study designs of comparative effectiveness, also known as, comparative effectiveness research or relative effectiveness assessment [33, 34]. In fact, the taxonomy of biases now surpasses 70 types extensively documented in the literature [35, 36]. The collection of data originating from an uncontrolled real life environment (i.e. monitoring study) cannot lead to the assertion of a causal relationship. This is because such a study design does not sufficiently isolate the effect of the treatment from the effect of the confounders. As such, it is imperative to use ex ante and/or ex post micro-econometric techniques like propensity score matching, instrumental variables, and adjustments by multivariate analysis [37-45] in order to ensure that the populations under study are comparable.

#### **4.4.2 Analytic techniques for addressing measured and unmeasured confounding**

Impact studies are designed to identify a cause-effect relationship by singling out the causal effect (i.e. impact) of a treatment. They attempt to correct most of the bias described above through the implementation of appropriate ex post micro-econometric measures. Below, we discuss the basic characteristics of suitable impact studies.

As in traditional studies, there is a test group and a control group. To be comparable, members of different groups must share the observable individual characteristics that determine the predicted probability of receiving the treatment or the likelihood of a doctor prescribing it to them. One of

these expressions is used to generate a one-dimensional score, otherwise known as the propensity score, with a logistic regression that combines all the known covariates.

When designing the study schematics, there are four steps that should be followed:

- First, an exposure model should be conceived independently of the expected treatment effects. The probability of getting the treatment is estimated in advance based on observable characteristics.
- Second, individuals in the treatment group are matched with those with similar propensity scores<sup>6</sup> in the control group. This makes the treatment assignment random, or at least independent of the outcome. The risk of bias is therefore reduced, using means that are more methodologically modest compared to those used in controlled trials.
- Third, a cause-effect model is formulated in order to study the observed mean difference between treated and untreated participants that have been matched based on their propensity scores.
- Finally, the double difference technique is implemented to reduce the bias caused by non-observable factors. This technique attempts to correct for undetectable factors that may affect the evolution of treatment throughout time. It consists of collecting the data of interest before and after the treatment, in a homothetic manner and across both groups. The before-after difference is estimated in both groups and the difference estimated in the control group is subtracted from the one estimated in the experimental group. Thus, neutralizing most unobservable time-bound third factors.

---

<sup>6</sup> Relying on the assumption that the groups' density distributions of those scores have at least some partial similarity.

In this context, the propensity score emerges as an ingenious strategy to balance the systematic differences between treated and untreated subgroups. Impact studies that are designed in this way enable the analyst to move beyond descriptive analyses and towards establishing inference by clarifying whether signals and tendencies are the result of confounded happenstance or in fact the effect of treatment. As such, impact studies can identify the true causes of observed phenomena with greater precision.

Whether the nature of the agreements are founded on health outcomes (i.e. pay for performance and CED) or financial commitments (i.e. price/volume agreements), or on the study design (i.e. descriptive, normative, quasi-experimental and experimental), an information system has to be continuously monitored via process and outcome indicators. Conversely, impact studies necessitate a real comparator in order to be implemented. So that these studies do not resemble a black box, the background parameters originating from monitoring studies need to be re-introduced in quasi-experimental designs as a way to permit propensity score analysis.

INSERT FIGURE 2

#### 4.5 Health outcomes vs. Performance *with reference to health outcomes*

Let us recapitulate the steps that constitute the causality chain: *means* are mobilized to support *activities* whose *products* (i.e. number of patients) result in *effects* (i.e. number of responders to treatment) measured by a set of ‘performance’ indicators. A number of externally defined *intermediary or process variables* are used to demonstrate that modifications effectively contribute to improving an individual’s health and/or lengthening of her/his life. This is the *ultimate impact*



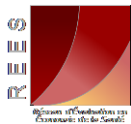
of the performance chain - the one and only true result in the patient's eye. A health service has only one purport *in fine*: its clinical utility. No matter what term is used to define the *clinical endpoint, impact, result, or clinical utility*, it all comes down to comparing the effect of two different treatments and estimating the *added value* of the innovation.

Under no circumstance can a monitoring study aid in understanding real treatment effects. The term 'effect' implies 'consequence of a cause'. This implies causality and it is unverifiable by a monitoring study. However, a monitoring study can evaluate the performance of a treatment. Performance, in this sense is but the observed consequence following treatment administration independently of whether or not the treatment is the actual cause.

A study may be able to compare the performance of two treatments, but that is a weak intermediary judgment. There is no certainty about the comparative effectiveness and the cost-effectiveness of the treatments in real life. As such, one-arm multicentric prospective studies based on non-validated intermediate endpoints as substitution criteria are average at best when compared to effectiveness studies. The latter are patient-centered and fundamentally impact studies.

INSERT TABLE 1

In sum, performance can be reduced to the rather administrative task of assessing whether external reference are met. This implies that a closer look at the actual magnitude and significance of the obtained health outcome is foregone. Measuring performance without regard for the real value of the innovation in question is a mistaken scientific endeavor and a blinded response to the demands of evidence-based policy-making. As an indicator, performance merely assesses whether a recognized process marker has been dully attained but does not provide insight on the real health



effect of the innovation. As such, performance measures offer a set of informed assumptions *with reference to* health outcomes, but *not directly or exclusively linked to* health outcomes. There is a tectonic difference, if only subtly misplaced by vocabulary. Should we be interested in applying econometric rigor to performance, one may realize that the attainment of “targets” bears no reflection on the certainty that they are in fact the direct effect of an innovation. For this, performance should not serve as a proxy for results<sup>7</sup>.

---

<sup>7</sup> This can be further exemplified in the work of Holloway et al. “Development of Performance measures for Acute Ischemic Stroke” (2001).

## 5 The way forward

The LEEM-CEPS Framework Agreements signed on December 5<sup>th</sup>, 2012 do not privilege any of the two possible studies over the other. The article 10<sup>ter</sup> of the agreement allows price-setting based on results observed in real life while using indicators conventionally decided upon with the manufacturer in question. Moreover, the methods to evaluate the effects of a new drug in real life are not chosen beforehand, the HAS plays a key role in the decision. The agreement states the methodological contribution of the HAS in article 11 paragraph 7: ‘before the preparation of the protocol, the HAS and the CEPS must agree upon the aims of the study and, consequently, the topics to be addressed.’ Paragraph 8 of the same article specifies that ‘the study protocol is submitted to the HAS for confirmation that the study is able to adequately address the issues of interest.’ Finally, it is indicated in paragraph 9 that the ‘HAS evaluates the results and determines with the CEPS to what extent the objectives have been met.’ The two institutions will, hence, collaborate narrowly.

## 6 Conclusion

We advance that performance-based risk sharing agreements present a risky move for both the payer and the manufacturer. Commitments are made about the performance of a treatment, but not about the value of that treatment. As long as the link between the intermediate and the clinical endpoints is not demonstrated through an impact study, the estimates of ‘performance’ obtained through registries have no scientific basis. It is a mere commercial proposition that allows to end dead lock negotiations without contributing in any way to evaluating treatment effectiveness in real life. While an impact evaluation of a treatment, verifies its added value in terms of clinical effectiveness, a performance-based evaluation is better suited for normative purposes. This is because the latter is a comparison of arbitrarily fixed values.

Our classification goes beyond that of Jaroslowski et al. [46] but reaches the same conclusion that neither ‘performance’ contracts with spending records and results obligations nor financial agreements bring any truly new elements relevant to the decision of whether or not the treatment is added to the list of ‘recognized expenses’ by the insurance firm. Such a decision can only be made in light of results originating from impact studies, with coverage for evidence development or post-inscription comparative studies.

Research on risk sharing contracts is at its beginning stages. There have been numerous articles published on the topic, but the majority of them regard descriptive work. Risk sharing contracts will become more popular in the future as an instrument that facilitates access to new treatments and enables patients to optimize their chances of survival. Notwithstanding, there is an emerging need to deepen methodological research on this topic in order to enable our discipline to adequately respond to the demands of the health industry and the public authorities.

## 7 References

- [1] Eunethta. Levels of evidence - Applicability of evidence in the context of a relative effectiveness assessment of pharmaceuticals: European Network for Health Technology Assessment2012. [Accessed December 19, 2013] Available from: <http://www.eunethta.eu/outputs/methodological-guideline-rea-pharmaceuticals-applicability-evidence-context-rea>.
- [2] Garrison LP, Jr., Towse A, Briggs A, de Pouvourville G, Grueger J, Mohr PE, et al. Performance-based risk-sharing arrangements-good practices for design, implementation, and evaluation: report of the ISPOR good practices for performance-based risk-sharing arrangements task force. *Value in Health*. 2013 Jul-Aug;16(5):703-19.
- [3] European Commission. Proposal for a Directive of the european parliament and of the council relating to the transparency of measures regulating the prices of medicinal products for human use and their inclusion in the scope of public health insurance systems. Brussels2012. [Accessed December 19, 2013] Available from: [http://ec.europa.eu/enterprise/sectors/healthcare/files/docs/transpadir\\_finalprop01032012\\_en.pdf](http://ec.europa.eu/enterprise/sectors/healthcare/files/docs/transpadir_finalprop01032012_en.pdf).
- [4] US General Accounting Office. Designing evaluations. Washington1991. [Accessed December 19, 2013] Available from: <http://www.gao.gov/products/GAO-12-208G>.
- [5] Neumann PJ, Chambers JD, Simon F, Meckley LM. Risk-sharing arrangements that link payment for drugs to health outcomes are proving hard to implement. *Health Aff (Millwood)*. 2011 Dec;30(12):2329-37.
- [6] Chapman S, Reeve E, Rajaratnam G, Neary R. Setting up an outcomes guarantee for pharmaceuticals: new approach to risk sharing in primary care. *BMJ*. 2003 Mar 29;326(7391):707-9.
- [7] Australian Government - Departement of Health and Ageing. Guidelines for deeds of agreement for the pharmaceutical benefits scheme. Camberra2009 Contract No.: 1.3. [Accessed December 19, 2013] Available from: <http://www.pbs.gov.au/info/industry/listing/elements/deeds-agreement>.
- [8] Menon D, Stafinski T, Nardalli AA, McCabe C. Access with Evidence Development Schemes: A Framework for Description and Evaluation. 2010.
- [9] Carlson JJ, Sullivan SD, Garrison LP, Neumann PJ, Veenstra DL. Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. *Health Policy*. 2010 Aug;96(3):179-90.
- [10] de Pouvourville G, Mongrédien L. L'accès au marché remboursé pour les médicaments: Les contrats de partage de risque fondés sur les résultats. Cergy-Pontoise, Le Kremlin-Bicêtre: ESSEC Business School, Collège des Economistes de la Santé2012. [Accessed December 19, 2013] Available from: [http://www.ces-asso.org/sites/default/files/Contrats\\_de\\_partage\\_des\\_risques.pdf](http://www.ces-asso.org/sites/default/files/Contrats_de_partage_des_risques.pdf).
- [11] Adamski J, Godman B, Ofierska-Sujkowska G, Osinska B, Herholz H, Wendykowska K, et al. Risk sharing arrangements for pharmaceuticals: potential considerations and recommendations for European payers. *BMC Health Serv Res*. 2010;10:153.
- [12] Rochaix L. Asymétries d'information et incertitude en santé : les apports de la théorie des contrats. *Economie & prévision*. 1997;129(129-130):11-24.
- [13] Carbonneil C, Quentin F, Lee-Robin SH. A common policy framework for evidence generation on promising health technologies. *Int J Technol Assess Health Care*. 2009 Dec;25 Suppl 2:56-67.
- [14] Ministère en charge de la santé. Prise en charge de l'innovation par la DGOS : de la recherche clinique à la diffusion aux patients. Paris2012 [December 19, 2013]; Available from:

<http://www.sante.gouv.fr/prise-en-charge-de-l-innovation-par-la-dgos-de-la-recherche-clinique-a-la-diffusion-aux-patients.html>.

- [15] Ranque B, Mechtouff L, Grabar S. Épidémiologie étiologique : du facteur de risque à la cause. Sang Thrombose Vaisseaux. 2011;23(5):242-52.
- [16] Haute Autorité de Santé. Choix méthodologiques pour l'évaluation économique à la HAS. Paris2011. [Accessed December 19, 2013] Available from: [http://www.has-sante.fr/portail/upload/docs/application/pdf/2011-11/guide\\_methodo\\_vf.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/2011-11/guide_methodo_vf.pdf).
- [17] Becquemont L, Bordet R, Cellier D. La médecine personnalisée : comment passer du concept à l'intégration dans un plan de développement clinique en vue d'une AMM ? Therapie. 2012 Jul-Aug;67(4):339-48.
- [18] Office parlementaire d'évaluation des choix scientifiques et technologiques. Etude de faisabilité de la saisine sur "les enjeux scientifiques, technologiques et éthiques de la médecine personnalisée". Paris: Assemblée Nationale; 2013 [December 19, 2013]; Available from: <http://www.assemblee-nationale.fr/opecest/ba14.asp>.
- [19] Institut national du cancer. Situation de la chimiothérapie des cancers. Rapport 2012. Paris: Institut national du cancer2013. [Accessed Available from].
- [20] Misdrahi D, Delgado A, Bouju S, Comet D, Chiariny JF. [Rationale for the use of long-acting injectable risperidone: a survey of French psychiatrists]. Encephale. 2013 May;39 Suppl 1:S8-14.
- [21] Bourassa Forcier M, Noël N. Ententes entre gouvernements et compagnies pharmaceutiques. Montreal: Centre interuniversitaire de recherche en analyse des organisations; 2012 [December 19, 2013]; Available from: <http://econpapers.repec.org/paper/circirpro/2012rp-13.htm>.
- [22] Comité économique des produits de santé. Rapport d'activité 2011. Paris2012. [Accessed December 19, 2013] Available from: [http://www.sante.gouv.fr/IMG/pdf/Rapport\\_Activite\\_CEPS\\_Final\\_2011-2.pdf](http://www.sante.gouv.fr/IMG/pdf/Rapport_Activite_CEPS_Final_2011-2.pdf).
- [23] Khandker SK, GB; Samad, HA. Handbook on Impact Evaluation: Quantitative Methods and Practices. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2009.
- [24] Gertler PM, S; Premand, P; Rawlings, LB; Vermeersch, CMJ. Impact Evaluation in Practice. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2010.
- [25] Savedoff WDL, R.; and Birdsall, N. . When will we ever learn? Improving Lives through Impact Evaluation. Washington, DC: Center for Global Development2006. [Accessed December 19, 2013] Available from: [http://international.cgdev.org/files/7973\\_file\\_WillWeEverLearn.pdf](http://international.cgdev.org/files/7973_file_WillWeEverLearn.pdf).
- [26] Gliklich RE, Dreyer NA, editors. Registries for Evaluating Patient Outcomes: A User's Guide. Rockville: Agency for Healthcare Research and Quality; 2010.
- [27] Eichler R. Utilisation des paiements basés sur la performance pour améliorer les programmes de santé. Le Management. 2001;10(2):1-22.
- [28] Circulaire DSS/1C/DGOS/PF2 no 2010-389 du 12 novembre 2010 relative à la mise en oeuvre des dispositions relatives à la maîtrise des produits de santé des listes en sus/actions locales à conduire en 2010 et 2011 (application du dispositif de régulation), (2010).
- [29] Contandriopoulos A-P, Champagne F, Denis J-L, Pineault R. L'évaluation dans le domaine de la santé: Concepts et Méthodes. Bulletin. 1993;33(1):12-7.
- [30] Gridchyna I. Utilisation de la norme juridique comme instrument de régulation du marché des médicaments innovants en Europe et en France. Bordeaux: Université Bordeaux 2; 2012.
- [31] Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational psychology. 1974;66(5):688-701.
- [32] Duflo EG, R; Kremer, M. Using Randomization in Development Economics Research: A Toolkit. MIT Department of Economics Working Paper. 2006(No 06-36).

- [33] Patient-Centered Outcomes Research Institute. Our Questions, Our Decisions : Standards for Patient-Centered Outcomes Research 2012. [Accessed December 19, 2013] Available from: <http://www.pcori.org/assets/MethodologyReport-Comment.pdf>.
- [34] European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. ENCePP Guide on methodological standards in pharmacoepidemiology. London 2010. [Accessed December 19, 2013] Available from: [http://www.encepp.eu/standards\\_and\\_guidances/documents/ENCePPGuideofMethStandardsinPE.pdf](http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE.pdf).
- [35] Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health. 2004 Aug;58(8):635-41.
- [36] Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32(1-2):51-63.
- [37] Givord P. Méthodes économétriques pour l'évaluation de politiques publiques. Paris: INSEE 2010. [Accessed December 19, 2013] Available from: [http://www.crest.fr/ckfinder/userfiles/files/Pageperso/givord/eval/slide\\_intro\\_Dauph\\_i.pdf](http://www.crest.fr/ckfinder/userfiles/files/Pageperso/givord/eval/slide_intro_Dauph_i.pdf).
- [38] Behaghel L. Lire l'économétrie. La Découverte, editor. Paris 2012.
- [39] Rosenbaum P. Randomized experiments and observational studies : causal inference in statistics. 2002.
- [40] Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. Journal of the American Statistical Association. 1984;79(387):516-24.
- [41] Denis P, Launois R, Devaux M, Berdeaux G. Comparison of diurnal intraocular pressure control by latanoprost versus travoprost : results of an observational survey. Clin Drug Investig. 2006;26(12):703-14.
- [42] Payet S, Riou Franca L, Le Lay K, Vallet B, Dhainaut JF, Launois R. Evaluation coût-efficacité de la drotrécogine alfa comparée à la prise en charge conventionnelle dans le traitement du sepsis sévère en pratique réelle. Journal d'Economie Médicale. 2007;25(4):207-23.
- [43] Dhainaut JF, Payet S, Vallet B, Franca LR, Annane D, Bollaert PE, et al. Cost-effectiveness of activated protein C in real-life clinical practice. Crit Care. 2007;11(5):R99.
- [44] Riou Franca L, Payet S, Le Lay K, Launois R. Drotrecogin alfa's impact on intensive care workload in real life practice: a propensity score approach. Value in Health. 2008 Dec;11(7):1051-60.
- [45] Grootendorst P. A review of instrumental variables estimation in the applied health sciences. Health Serv Outcomes Res Methods. 2007;7(3-4):159-79.
- [46] Jaroslowski S, Toumi M. Market Access Agreements for pharmaceuticals in Europe: diversity of approaches and underlying concepts. BMC Health Serv Res. 2011;11:259.

## 8 Table and figure captions

**Figure 1:** Classification of market entry agreements [9]

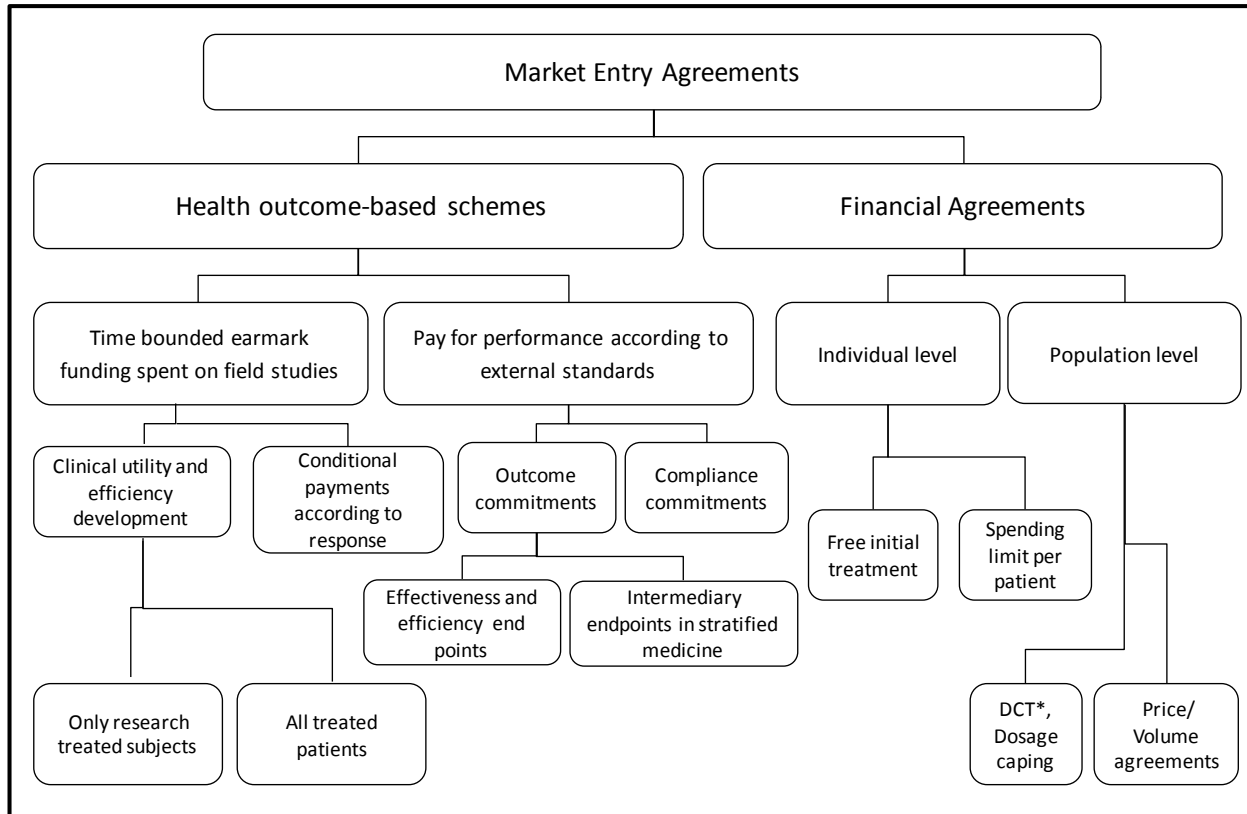
**Figure 2:** New way to evaluate “promising” treatments

**Table 1:** Conceptual setting of risk sharing agreements



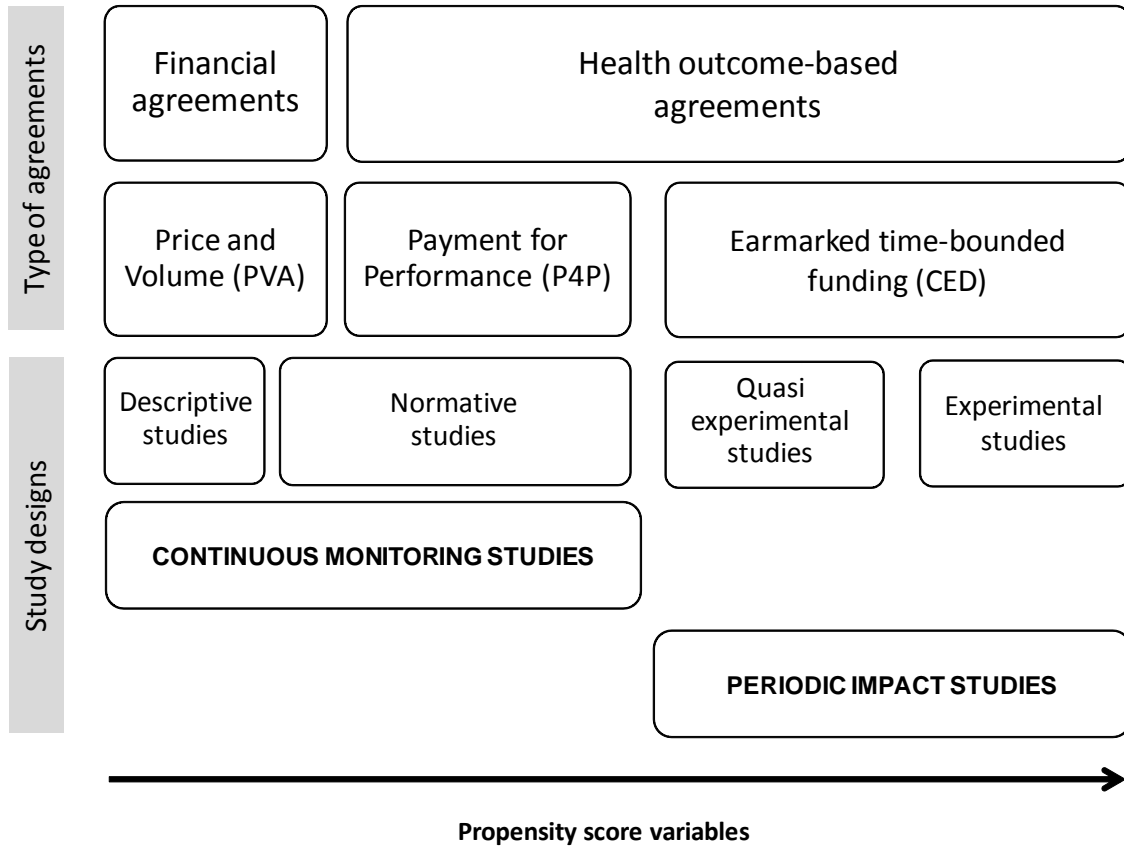
## 9 Tables and figures

**Figure 1**



Carlson, Garrisson et al 2010. \*DCT: Daily Cost of Treatment.

**Figure 2**



**Table 1**

<b>Obligations</b>	<b>Health outcome-based agreements</b>		<b>Financial agreements</b>
<b>Nature of the Study</b>	<b>Impact studies</b>	<b>Monitoring studies</b>	
<b>Designs</b>	Comparative studies ECR/CEA†	Audit/product registry/ DUR/P4P‡	Audit/product registry /MADB BIM/PVA*
<b>Indicators</b>	Clinical utility and efficiency	Response predicting marker in a stratified medicine context	Target values
<b>Specific clauses</b>	Refund available for study participants or for all eligible patients	Reimbursement if no response in certain groups or biomarkers T- Different prices according to sub-groups	Spending limit per patient, monitor dosing, price/volume agreements
<b>Endpoints</b>	Limitation de l'incertitude sur l'utilité clinique & médico-éco	limiting results uncertainty within groups & respect guidelines	Limit budgetary uncertainty

† CER : Comparative Effectiveness Research; CEA: Cost Effectiveness Analysis; ‡ MADB : Medico-Administrative Databases; DUR: Drug Utilization Reviews; P4P : Pay for Performance ; \* BIM : Budget Impact Model; PVA : Price/Volume agreements