Chapter 3
The Scope of Health Insurance

3.1 Measuring the Degree of Protection

3.1.1 Defining Quantity

In a conventional market, quantity refers to the amount of a particular good or service that a consumer receives from a producer. In the simplest models of supply and demand, it is easiest to think of simple, universal goods, like food, or even abstract goods that do not exist (e.g., the infamous economic "widget"). Then, this allows economics to conceive, and explain, the amount of a particular good that a person consumes at a particular price—how many apples does someone eat, or how many gallons of gas does someone use? There is also an underlying assumption here of substitutability of commodities. Any given gallon of gas is completely fungible to a driver, meaning that two gallons of gasoline with the same quality rating are identical.

The most general model of insurance looks at the amount of protection that people would choose in order to protect their portfolio as a whole. Health capital is part of the broader set of assets that individuals hold in their portfolio representing their overall wealth. Risk averse individuals seek to insure part or all of their wealth in order to avoid the losses associate with wealth shocks. Gollier (2001) explains the "standard portfolio problem" in the following way: "… a risk-averse agent who owns an asset which is subject to a random loss $\tilde{y}$… can select the share of the risk α that he will retain. The proportion $(1 - \alpha)$ is sold to an insurance company against the payment of an insurance premium that is proportional to the expected indemnity, which is also called the actuarial value of the policy" (Gollier 2001, p. 61).

In the model of the standard portfolio problem, individuals choose the level of risk they wish to retain in order to trade-off their desire to reduce their risk (risk aversion) against the cost of insurance. This model is an excellent way to conceptualize insurance for health capital, since people are literally endowed at birth with their health capital and are always fully at risk for health shocks and

R.D. Lieberthal, *What Is Health Insurance (Good) For?*,
DOI 10.1007/978-3-319-43796-5_3

depreciation unless they choose to obtain health insurance. The term $(1 - \alpha)$ is the measure of the overall amount or quantity of protection that is obtained in the insurance contract, and the term α is the measure of the amount of risk an individual chooses to retain. When $\alpha = 0$, $(1 - \alpha) = 1$, and the individual has full insurance. When $\alpha = 1$, $(1 - \alpha) = 0$, and the individual is "uninsured" or "self-insured." In this framework, risk protection is a pure commodity in the sense that two insurance contracts that each offers the same amount of protection to a given individual are considered as identical by that individual.

Both the individual and the insurer face the question of how much risk to manage in terms of setting the appropriate α. When purchasing health insurance, the individual must decide how much risk to retain. The insurer faces the opposite question—how much risk does it want to take on? Individuals who accept an insurance arrangement with a higher α retain more of the risk related to their health capital. Insurers that set $(1 - \alpha)$ higher are taking on more of the risk related to an insured's health capital. For this reason, α is also referred to as the "coinsurance rate" that individuals accept. Individuals who purchase partial insurance are, in some sense, "coinsurers" along with their health insurance company. A person who obtains a health insurance contract for 80 % of their risk, i.e., with $(1 - \alpha) = 0.80$, has a coinsurance rate of 20 %, i.e., $\alpha = 0.20$. That individual retains 20 % of the risk under such a health insurance policy.

In practice, health insurers can split the risk for health capital with individuals in numerous ways. In addition to coinsurance, the main ways of limiting the scope of health insurance coverage are deductibles, copayments, and coverage limits. Each of these forms of individual financial responsibility corresponds to a lower $(1 - \alpha)$. In other words, these forms of financial responsibility correspond to a higher α, the amount of risk retained by the individual. However, copayments, coinsurance, deductibles, and coverage limits all have different implications for the financial responsibility of the individual. Coinsurance is based on a proportionate share of the risk, while copayments, deductibles, and coverage limits are based on a fixed dollar amount. Copayments are generally associated with services that have a smaller cost, whereas coinsurance is more important for higher cost services or individuals with higher expected healthcare costs. Copayments refer to individual claims or specific interventions, whereas deductibles and coverage limits pertain to the overall losses incurred by an individual over a longer period of time—a yearly insurance policy, or even over a lifetime. Table 3.1 shows examples of how these forms of individual financial responsibility can be applied.

Risk averse individuals also have certain preferences for different types of health insurance designs. From the point of view of the insured individual, each form of financial responsibility has different benefits and costs. This is the same for the insurer. Coinsurance has a value to individuals in that it implies a payoff for even the smallest claims. Coinsurance reduces the price of all healthcare services somewhat like a coupon—a 20 % coinsurance contract is equivalent to "80 % off" all healthcare services. Deductibles have a value to individuals in that it limits coverage for only the smallest losses, thus concentrating health insurance payouts in the more catastrophic situations. Under deductibles, individuals have certainty

Table 3.1 Limitations on health insurance coverage

Limitation on health insurance	Form of measurement	Examples of application
Coinsurance	Proportional	The individual pays 20 % of each claim
Deductible	Absolute	The individual pays the first $1000 of each claim The individual pays the first $5000 of claims incurred in a year
Copayment	Absolute	The individual pays $10 for each doctor's visit
Coverage limits	Absolute	The individual pays for all costs for a claim exceeding $1 million The individual pays for all claims costs exceeding $3 million in a given year The insurer's payout to an individual over their lifetime is capped at $5 million

about their maximum out-of-pocket costs. Copayments lower the premium for insurance for a relatively small cost. A $10 copayment for a doctor's visit is not onerous for people with financial means, and so they may be willing to accept such copayments in order to obtain lower cost insurance. Coverage limits are mainly beneficial for reducing the cost of insurance, since an insurer will charge lower premiums to individuals who are willing to limit their coverage from insurance.

The mix of forms of financial responsibility and the expected claims under the policy are used to measure the quantity of protection offered by a health insurance contract at the time the contract is written. This calculation is also known as "actuarial value," or the percent of total expected healthcare costs that a particular plan is expected to cover within a given year. For example, if an individual is expected to incur $5000 in healthcare expenses in a given year, and his or her plan is expected to pay $4000 of those costs, then it has an actuarial value of 80 %. This is equivalent to a plan with a coinsurance rate of 20 %. The actuarial value does not correspond precisely with coinsurance levels, since a plan can achieve a particular value level through a combination of coinsurance, deductibles, copayments, and coverage limits. For example, insurance policy A with a 20 % coinsurance rate and an annual deductible of $5000 might have an $\alpha = 0.25$ for a particular individual. If insurance policy B with a 15 % coinsurance rate and a $15,000 annual deductible has $\alpha = 0.30$, then policy A actually provides a greater degree of protection than policy B, despite the fact that policy A has a higher coinsurance rate. This calculation is also known as an *ex ante* calculation, because it is based on the expected claims under a contract rather than the actual claims.

It is also possible to measure the quantity of protection on an *ex post* basis. After an insurance contract ends, an individual can measure the total out-of-pocket payments and total healthcare spending in order to determine of how much of their healthcare spending was covered by health insurance. On average, this amount should be equal to the *ex ante* measurement of quantity, but individual experience will vary substantially. For example, an individual who purchases a policy with a $1000 deductible and has claims of only $500 during a plan year would receive no

financial payments from the insurance company, which is equivalent on an ex-post basis to being uninsured, i.e., $\alpha = 1$. Conversely, an individual with a $1000 deductible who experiences an expensive injury costing $300,000 will have almost all of their his or her claims paid for by the insurance company, which is nearly the same as having had full insurance, i.e., $\alpha = 0$. The ex post calculation of payments is also used to calculate the "medical loss ratio" of an insurance contract, meaning the proportion of premiums spent on medical claims.

An important aspect of quantity also relates to the amount of financial intermediation an individual receives under a health insurance policy. There are degrees of intensity of third-party payment. Indemnity insurance, where an individual receives an insurance settlement for a loss as a lump sum from the insurer could be said to receive relatively less in third-party payment services. An insurer who makes the payment to the provider directly on a service basis could be said to be providing a moderate amount of third-party payment services. There are also larger quantities of third-party payment, such as are provided by managed care payers. For example, the insurer that bundles all payments for an episode of care into a single amount, or that pays for healthcare on a capitated basis, could be said to be providing a large amount of third-party payment services.

3.1.2 Measuring Quality

While the measurement of quantity of insurance is used to examine health insurance as a commodity, substantial variation exists in the way that health insurers apply the same health insurance contract. This variation relates to the subjectivity of health capital—after all, the purpose of insurance is to relate the percentage of a loss suffered to the amount of payments made by the individual and insurer. If the amount of the loss itself is measured with some degree of error, then two insurance plans with a 20 % coinsurance rate on the same individual could pay out different amounts based on a different valuation of that individual's human capital. Higher quality insurance minimizes this error.

Quality of health insurance can be measured from the point of view of the consumer or the provider. The reason to measure the quality of coverage from the consumer point of view is that the motivation for having health insurance ultimately comes from the risk averse individual who wishes to protect their health capital. The quality of health insurance is ultimately one of the determinants of how much utility that individual derives from the health insurance arrangement. From the point of view of providers, the reason to measure health insurance quality is that their monetary incentives determine the quantity and quality of healthcare that they provide. Providers also possess a key informational advantage relative to consumers in that they have greater information about variation in health plan quality. This advantage becomes particularly important with the measurement of the quality of health insurance, since this is a less tangible dimension for assessing health insurance than the quantity.

In regards to health plan quality, consumers care about the amount of latitude they have in terms of selecting a provider, also known as the breadth or narrowness of networks. Consumers care because of differences in provider skills and outcomes with certain conditions, as well as because of more humanistic preferences, i.e., "bedside manner." Consumers may want the provider with the most expertise in their particular condition because that provider will do a better job of helping them to maintain their health capital. Insurers can choose to contract with a limited number of providers in order to reduce the cost of health insurance, potentially leading to differences in the quality of insurance across plans with the same quantity of financial protection.

The quality of insurance can relate to large price differences faced by individuals for services and health insurance premiums. For example, an individual seeking a knee replacement surgery may face a large difference in the price for healthcare between "in network" and "out of network" surgeons. Narrow networks lower the perceived quality of the insurance for the consumer to the extent that an out-of-network surgeon has greater experience, has better outcomes of care, or is an individual's preferred provider. Providers share this view of quality with consumers, in the sense that they may perceive an insurer that pays them a higher reimbursement as being of higher quality. Providers are more likely to accept forms of insurance that pay more, such as private health insurance, and less likely to provide care to patients whose insurance pays lower rates such as Medicaid (Decker 2012). Insurers could use these preferences to command a higher premium for insurance with higher perceived quality, holding quantity constant.

Providers also view quality in terms of the strictness of claims reviews. Insurers perform a crucial third-party payment role in terms of structuring reimbursement. They then apply those rules to any claims incurred under the insurance arrangement. Higher quality insurers from the provider point of view would make the process of claims review easier, in the sense that an honest provider submitting what they consider to be a legitimate claim is required to submit a reasonable amount of documentation. Providers may consider insurers as lower quality if they require a large amount of documentation, prior authorization of specific procedures or drugs, or other forms of "utilization review."[1]

Providers may also have preferences about the speed of claims payments. The time spent waiting for claims is costly for providers and facilities, because they must find a way to finance the cost of their operations while waiting for accounts receivable to be paid. There is some uncertainty over whether, and when, insurers will pay claims because the insurer acts as an intermediary, adjudicating and paying claims.

[1]The literature on physician and provider perception of health insurance quality is less developed than the literature on patient and consumer perception and choice of health insurance. Prior studies have found that physicians have a strong opinion that health insurers, among others, "… have a 'major responsibility' for reducing healthcare costs …". The same study, when examining physician's opinions about third-party payment, found that "Few expressed enthusiasm for 'eliminating fee-for-service payment models' (7 %)" (Tilburt et al. 2013). Future research into provider perceptions of health insurance may be a fruitful area for future research.

However, providers cannot "repossess" improvements in health capital in the same way that is possible with a physical asset such as a car, so they take on some risk when they provide care in exchange for a future payment from an insurer. Higher quality insurance pays the same claims more quickly, lowering financial risk for providers.

Health services research is largely concerned with measuring the quality of health insurance through health outcomes. Higher quality insurance does better job of improving health outcomes such as survival and quality of life. To the extent that a causal link can be established between the provision of health insurance and greater health, the insurance can be said to be of higher quality. For example, there is some evidence that transitioning from being uninsured to insured can reduce mortality. For instance, in studying the effects of the Massachusetts universal insurance reform of 2010 extended health insurance to previously uninsured individuals, Sommers and colleagues found that the insurance expansion reduced mortality rates for those who were most exposed to the shift from uninsured to insured (Sommers et al. 2012, 2014).

While mortality is an objective measure of health insurance quality, the evidence base linking better insurance to lower mortality is limited. As a result, it is difficult to use mortality as a proxy measure for quality. In many cases, studies that have examined this question find no causal link between health insurance quantity or quality and mortality. Canonical studies including the RAND Health Insurance Experiment and the Oregon Medicaid Experiment showed improvements in measures like blood pressure and depression, but little or no connection between more insurance and lower mortality (Baicker et al. 2013; Manning et al. 1987).

Health services research also suggests ways that health insurance can directly address the basis risk inherent in health insurance. Health insurance plans that utilize a "value-based insurance design" attempt to adjust measures of financial responsibility to reflect the underlying effectiveness of care in different populations (Chernew et al. 2007). Insurers that utilize "risk based contracting" attempt to make payments to healthcare suppliers contingent on evidence about outcomes (Frank et al. 1995). However, little evidence exists for a wide range of conditions. There are also many conditions for which the majority of outcomes are more subjective, or based on patient reporting, especially pain and mental illness. Thus, there is a multitude of ways to measure the quality of insurance through objective and subjective health outcomes.

In many cases, it is possible to achieve a lower standard or evidence and show that higher quality insurance ensures more "access" to healthcare. Access to healthcare relates to the measures of narrowness of networks, claims review, and timeliness of payments. The financial way to measure access is to examine the extent to which the presence of insurance impacts the overall consumption of healthcare or out-of-pocket payments. However, financial considerations are only one of the potential obstacles to care—location, supply of providers and facilities, and individual engagement all play a role in access, as well, so access is often measured either in terms of the quantity of care obtained ("utilization") and in terms of health outcomes related to care (Gold 1998). Access also has a specific dimension related to particular types of health shocks. To what extent is an individual who suffers a heart attack able to obtain acute care that directly addresses this

specific condition? To what extent is this individual able to obtain follow-up care from a cardiologist or a primary care doctor in order to monitor their condition and plan for prevention of future heart attacks? Proxy measures for access generally rely on supply or wait times, with the idea that it will be quicker and easier for a person with higher quality insurance to find a provider or facility that is willing to provide treatment (Thompson et al. 1998).

Higher quality insurance also has better customer service. A plan that is more responsive to its members is of higher quality in much the same way that a plan that pays claims faster is of higher quality. A number of measures have been developed to assess the responsiveness of health insurance and healthcare more generally. The Healthcare Effectiveness Data and Information Set (HEDIS) measures provided by NCQA and the Medicare "star" rankings of Medicare Advantage plans that may help individuals select a health plan are two examples of quality rankings based on customer service or satisfaction (Reid et al. 2013; Thompson et al. 1998). As with other measures of quality, these customer service rankings are somewhat intangible, and may even be uncorrelated with, or inversely correlated with, other measures of quality and quantity.

One other important aspect of health insurance quality is the term of protection. The term of protection refers to the amount of time that an insurance contract covers an individual. For example, insurance that covers all claims for a given year has a one-year term. In many countries with universal health insurance, the term of protection is equal to a person's lifetime. In the U.S., most health insurance is written on an annual basis (one-year term). That means that any decreases in health that occur within a particular year are paid for by the health insurer. It also means that, at the end of the plan year, the individual with insurance faces the possibility that their premium will change to reflect any changes in their expected claims in the future.

A defined term has the effect of splitting coverage for decreases in health into one-year segments. Annual renewal coverage works best for discrete, random health shocks of the kind that a risk averse individual is most concerned about. For example, the individual who has a heart attack without a previously identified cardiac problem has suffered a discrete health shock. The insurer's job as a risk manager is to anticipate, and pay for, the care of that proportion of the insured population that suffers a heart attack. One-year segments are less appropriate for longer term depreciations in health. For example, diabetes is a chronic disease with an effect that takes many years to accumulate. It is difficult, if not impossible, to identify "the day" that a person contracted diabetes. Instead, markers that would indicate that an individual may have diabetes increase over time (Tabák et al. 2009). Under a system of one-year term insurance, a person with diabetes would be charged an actuarial premium that incorporates the probability of needing care within a certain year. For an individual with a disease like diabetes, that cost is likely to be higher than for the non-diabetic population.

Alternatively, an insurer could exclude coverage for certain types of health conditions, also known as "pre-existing condition" exclusions. However, such exclusions are no longer allowed in the U.S. as a result of the ACA (Patient Protection and Affordable Care Act 2010). In the past, the use of pre-existing

condition exclusions meant that many individuals either could not obtain private health insurance or that the quantity of coverage was limited to "new" injuries or illnesses. Now that the practice is disallowed, more individuals may be able to obtain coverage, especially through the nongroup market; at a price they find to be reasonable. However, the extent of increased coverage may be limited by the unwillingness of insurers to write health insurance when they are not allow to include pre-existing condition exclusions in their insurance contracts.

The literature that investigates the empirical question of how long individuals retain health insurance is known as the "persistence of health insurance" literature. For example, one of the major criticisms of nongroup and small-group insurance is that individuals tend to drop coverage or alternate between being insured and being uninsured. For example, Pauly and Lieberthal (2008) find that "… becoming uninsured is most likely for those with individual insurance, less likely for those with small-group insurance, and least likely for those with large-group insurance." This persistence depends to an extent on health status: "However, for people in poor or fair health, the chances of losing coverage are much greater for people who had small-group insurance than for those who had individual insurance." Analyses have also found that "… lower income and education were associated with not gaining and with losing private insurance. Poorer health status was associated with public insurance gain" (Jerant et al. 2012). The ACA will likely increase the persistence of coverage by increasing Medicaid eligibility for those with lower incomes and providing subsidies for nongroup insurance coverage.

3.1.3 Determining Prices

In a market, price refers to the amount of money that a consumer gives and that a producer receives for a particular good or service. We can take the set of prices as given: "… the simplest kind of market behavior (is) that of **price-taking behavior**. Each firm will be assumed to take prices as given …" (Varian 1992, p. 25). In this chapter, we also assume that consumers are "price-takers" as well, in that they take health insurance prices as given. Then, at a particular price, generally expressed in money terms, the consumer can obtain a specified quantity of a particular good from a producer.[2] Price setting behavior is important later in the book, since a large enough insurers could affect prices through their behavior, especially if it were a monopolist. Similarly, a large enough health insurance purchaser could affect prices through their behavior, especially if it were a monopsonist.

[2]Money itself is not even required in these models. The basic mathematical technique involves what is called a "numeraire" good, or a good whose price and quantity is fixed in order to index the sets of prices and quantities of all other goods (Mas-Colell et al. 1995, p. 325). It is a convenience to consider money to be this good, since it is a unit of exchange with a fixed value, i.e., a dollar has a fixed value of one dollar. In an asset pricing model, it is convenient to think of the numeraire good as a risk-free asset, meaning one with a fixed return (Gollier 2001, p. 332).

In an insurance market, premium refers to the amount of money that a consumer gives and that a producer receives for insurance. The premium is composed of two basic elements. One is the expected claims for the insurance policy. The quantity of insurance provided, $(1 - \alpha)$, determines the claims cost of the insurance contract. The other element is the loading factor for insurance. Insurance is costly to create, underwrite, and market. In addition, the insurer will require capital and a profit margin. All of these additional costs are added to expected claims to calculate the total premium (see Bluhm 2007, pp. 146–148 for additional detail).

The price of the insurance when we consider it as a financial asset is "… its expected payoff plus a risk premium. This risk premium is measured by the covariance of the payoff of the risky asset with the price kernel" (Gollier 2001, pp. 332–333). However, some health economists define the price of health insurance as the loading factor alone. "The price is not simply the premium paid, because that premium includes the average expense of something the consumer would have to pay anyway. The price of insurance is just any markup above those expected benefits that the insurance company adds" (Phelps 2003, p. 330).

The insurer uses the features of the insurance arrangement to calculate a premium for the insurance depending on the expected claims under the contract. Under an "actuarially fair" insurance contract, the premiums are set equal to expected claims. Under an "actuarially unfair" insurance contract, the premiums exceed the expected claims, meaning that there is a positive loading factor for the insurance. Under an "actuarially favorable" insurance contract, the premiums are less than the expected claims, in that there is a negative loading factor for insurance.

Actuarial fairness relates to the price of risk management. An actuarially fair contract has a price of zero for risk management, while an actuarially unfair contract has a positive price for risk management. An actuarially favorable contract has a negative price for risk management—an individual with such a contract is being "paid" to relinquish their risk. The price of risk management also determines the willingness of individuals to purchase insurance in relation to their preferences about risk. Risk neutral individuals will only purchase actuarially fair or actuarially favorable insurance—they will not pay for risk management. Risk averse individuals are willing to purchase actuarially unfair insurance as long as they are not "too unfair." The reason is that risk averse individuals have a positive demand for risk management, and so are willing to pay a price for risk management that is positive as long as it is not "too high."

3.2 Optimizing Health Insurance

3.2.1 Health Insurance Trade-offs

The willingness to pay for insurance based on the varying risk aversion of individuals is important because of the trade-offs inherent in the purchase of health insurance. From a societal point of view, money spent on health insurance is not

spent on other goods and services. In addition, the variation in the quantity, quality, and price of insurance means that different health insurance arrangements have different associated trade-offs. Looking at health insurance as part of a larger budget set for consumption is also useful because of the nature of health capital as a complement to almost all other forms of consumption—how healthy a person is determines how much utility they derive from what they eat, where they live, and how they interact with others. Allowing for different individuals to obtain differing degrees of protection for health capital can, in theory, allow each individual to trade-off the degree of protection that they receive with the amount spent on that protection.

Health insurance trade-offs also apply to insurance purchased by employers and the government on behalf of employees and citizens, respectively. Employers that spend a dollar on health insurance must either pay that cost through reduced profits or transfer the cost to workers through lower cash wages (income). Governments that spend a dollar on health insurance must either pay for that spending through reduced spending on other programs, or through higher taxes on individuals in order to finance that spending.[3] Employers and governments face the same choice about how much to spend on health insurance relative to all other goods that is faced by individuals.

We can disentangle the optimal insurance problem into two smaller problems—whether to obtain insurance, and then how much insurance to obtain conditional on obtaining health insurance. The reason to split the insurance decision in this way relates to the insurance models introduced in Chap. 2, which demonstrated that some individuals may be better off with insurance while other individuals are better off without insurance. Economists describe this choice as the "extensive margin" between having insurance and being uninsured. In health insurance, the insured/uninsured decision is also important because health insurance choice is not continuous—individuals face restrictions on low quantity insurance policies where α is less than, but close to, one (Gruber 2008).

Conditional on obtaining health insurance, the next choice is along the "intensive margin" of how much insurance to purchase. The models of risk aversion generally view the intensive margin by examining slightly different amounts of insurance that an individual could purchase at different price points. An individual might choose between two insurance policies where $\alpha = 0.20$ and where $\alpha = 0.15$. The latter policy offers more protection at a higher premium, meaning less money to spend on alternatives for health or non-health consumption. The diversity of individuals with respect to their level of health, degree of risk aversion, and degree of financial resources means that the benefits and costs of insurance differ by individual. A less healthy individual, a more risk averse individual, and an individual with more financial wealth to protect are each better off with the greater quantity of insurance, all else equal.

[3]Governments could also borrow the money, but such borrowing simply passes on the cost to future taxpayers in the form of taxes to repay those bonds.

A person who has a desired level of health capital also must consider the full menu of options for protection of health capital when considering the relative value of health insurance. Three important alternatives to health insurance are healthcare, prevention, and precautionary savings. In other words, a dollar spent on health insurance could instead be spent directly on healthcare, thus achieving a more direct increase in health capital. Prevention is another alternative to health insurance that is designed to reduce the probability or severity of future health losses, thus lessening the need for health insurance. Precautionary savings describes a form of saving behavior whereby a person saves it in order to have a buffer against future health spending (Kazarosian 1997). The advantage of precautionary savings is that a dollar saved can be spent on anything in the future—a person who continues to be healthy can spend the money on non-healthcare consumption. Each of these three alternatives avoids the need to pay the loading costs associated with health insurance. The main advantage of health insurance relative to each of these choices is that it covers a much broader array of health risks, meaning a wider array of health shocks (contingencies).

Economics also looks at health insurance as a service that changes wealth in different states of the world as well as changing the price of healthcare. Health insurance gives the purchaser a payment that is contingent upon future states of poor health. Thus, health insurance in some sense "transfers" wealth from healthy to unhealthy states of the world, getting close to the notion of health insurance as a way to "save" good health today for use tomorrow. Health insurance changes the price of healthcare through the delivery of in-kind rather than cash benefits from health insurance. The delivery of in-kind benefits changes the individual's budget set so that healthcare is a relatively more attractive consumption good from a price point of view. Thus, another reason to obtain health insurance—rather than healthcare, prevention, or precautionary savings—is that health insurance delivers a lower price for healthcare precisely when a person wants it the most.

One other important drawback of health insurance is that it could potentially increase risk in that what it pays for and how much it pays is uncertain. The subjectivity of health capital valuation and of health insurance coverage means that the purchase of health insurance mitigates risk rather than eliminating it. This is particularly true given the fact that health insurance benefits come in terms of services rather than cash. Forward-looking individuals may be unsure whether the services they will receive in the wake of a decrease in health will match their preferences for the type of care they would like to receive. Even with "full insurance", there is some uncertainty over outcomes of care in that no insurance policy can restore individuals to the state of health that they were in before they suffered a loss. Finally, individuals with insurance are subject to the uncertainty related to the ability of providers to exercise their expert judgment about their condition and what "should" be covered, as well as potential disagreement between the insurer and the provider.

Given that the marginal cost of health insurance is not zero and the marginal benefit is not infinite, the value of health insurance is finite. Health insurance is costly in terms of structuring policies, determining which claims should be paid,

and paying those policies. It is also costly from the consumer point of view in terms of search and selection of the health insurance policy and the payment of premiums. The marginal benefit of health insurance is finite because the risks related to health are finite—a person's health capital does not have an infinite value from an economic point of view. In fact, there is likely a declining marginal benefit to improvements in health capital and increasing marginal costs. The economic implication is that there is likely an "optimal" amount of insurance that balances out marginal cost and marginal benefit (Pauly 2000).

3.2.2 Optimal Health Insurance as a Benchmark

Optimal health insurance is the economic answer to the question of what is best in health insurance. Optimal health insurance is the policy that balances the individual preferences for risk management and healthcare finance against the individual preferences for consumption of goods other than health insurance subject to their total budget for all consumption. In particular, optimal health insurance takes into account the value of health capital and the possibility that other activities—healthcare, prevention, and savings—could deliver improvements in health at a lower cost in many cases. Optimal health insurance is constructed by selecting a health insurance policy that equalizes the marginal benefit and marginal cost of insurance for a given individual. Setting marginal benefit equal to marginal cost is the general economic solution that maximizes the value of health insurance as part of a consumer's overall consumption bundle.

The equalization of marginal costs and marginal benefits implies that optimal insurance could be full, partial, or no insurance at all. For certain individuals, the optimal insurance policy may be full coverage. The reason that full coverage is an upper limit is because insurance for more than 100 % of a loss becomes speculation (see also the principle of "contribution" as described in Chap. 1). Full insurance also implies full consumption smoothing over healthcare spending, meaning that, no matter the degree of health, health spending is fixed at the amount of the health insurance premium paid. Full insurance could also be seen as a "global budget" for healthcare spending, in that the premium paid defines the total amount of spending for healthcare from a person's budget. This is attractive to a risk averse individual since it eliminates a source of financial risk that they face. For most individuals, the optimal insurance policy is partial insurance. Partial insurance is distinct from full insurance in that it does not cover the entirety of a loss. Almost all health insurance provided in health insurance markets is partial insurance that imposes coinsurance, copayments, deductibles, and other limitations on coverage for healthcare spending. For some individuals, remaining uninsured may be optimal, especially if that individual is risk neutral or risk seeking.

Even with partial health insurance, health insurance is in some sense more full than other forms of insurance in that it also allows consumers to pre-pay for healthcare that is somewhat or fully predictable. This is the essence of the financial

intermediation or service basis for the payment of claims view of what health insurance can provide to individuals. For example, health insurance typically covers annual wellness visits at no additional cost (such coverage is now required under the ACA). If everyone in a group of homogeneous individuals agreed that they each needed a single annual physical examination (check-up), then they could each pay for the check-up on their own; there would be little use in pooling their own money and then having the pool pay out for each of their purchases (the cost of the examination). In fact, we could consider this pooling somewhat wasteful, since it adds a layer of complexity with little benefit.

Optimal insurance is an economic concept that serves as the benchmark against which we measure both the current system and the effectiveness of health insurance policies. In this chapter, we are examining the overall scope of health insurance before considering how markets for health insurance are actually set up. Optimal health insurance provides a sense of how close or how far health insurance as it currently exists is from how good it could be. Optimal health insurance also serves as a way to judge the value of efforts to improve the current health insurance system—even if the system results in suboptimal health insurance for many individuals, economics requires a calculation of the cost from transitioning from the current system to a better system of health insurance. Then, that cost of transition can be compared to the benefit associate with moving from suboptimal to optimal health insurance, in order to determine whether such a policy change passes the "cost-benefit" test.

One way to examine the somewhat intangible concept of changing the insurance system in terms of the degree of protection is to use examples from the Affordable Care Act (ACA). In general, the ACA moved many health insurance policies closer to full coverage. It did so by eliminating coverage limitations like lifetime limits on coverage and by limiting the amount that people can pay out-of-pocket through copayments, coinsurance, and deductibles. It also made certain types of policies with low annual limits illegal, both directly and by mandating a certain degree of coverage of "essential health benefits" in health insurance policies (Centers for Medicare and Medicaid Services 2016). This change moved individuals more toward having full coverage, since insurers must manage a certain portion of risks. On the other hand, this may lead to some people electing not to purchase coverage, if they perceive that the premiums are too high. In order to judge the value of these policies, we need to understand how economics views the optimization of health insurance coverage, meaning the process by which the "optimal" amount of insurance is determined for individuals and for society.

3.2.3 Determining the Optimal Policy

The economic process for optimizing any form of consumption is to consider it first in isolation, and then to add constraints to bring the conceptual model of the economy closer to reality. Economics first considers what optimal health insurance

would look like if insurance were the only way to protect health capital. That process allows us to isolate the effect of individual variables on the degree of protection that is optimal. For risk averse individuals facing the choice of actuarially fair health insurance, the optimal amount of coverage is full insurance, i.e., $\alpha = 0$ (Mossin 1968). The reason is that the marginal benefit of health insurance is always positive—moving from a higher to a lower α always makes a risk averse individual better off, because they now have relatively more protection against risk. Conversely, the marginal cost of health insurance is zero—moving from a higher to a lower α does not impose any costs on that risk averse individual because there is no loading cost. We can see that any risk averse individual would choose to obtain insurance on the extensive margin, and would choose full insurance along the intensive margin.

If health insurance were actuarially unfair, then the optimal policy for a risk averse individual is partial insurance. In other words, if there is some cost associated with insurance, the optimal health insurance policy is one where $0 < \alpha \leq 1$. Two early studies demonstrated this result: Arrow (1963) and Pratt (1964). We can understand the optimality of partial insurance by considering the two sides of this inequality, $\alpha > 0$ and $\alpha \leq 1$. The reason that full insurance is not optimal relates to the price of risk being positive. Once the price of risk is positive, individuals must trade-off spending on insurance against spending on other goods, and therefore will choose not to purchase full insurance but rather to allocate part of their budget to insurance, and the rest for other spending. The reason that $\alpha \leq 1$ is the amount of insurance that is optimal will vary by individual. The higher the price charged by an insurer in terms of the loading cost, the less an individual will wish to allocate to health insurance. In the limit, a loading cost could be high enough that even a risk averse individual would find the insurance to be "too expensive" and thus set $\alpha = 1$, i.e., choose to remain uninsured.

Prior studies have used the optimality of partial insurance to motivate the use of coinsurance, deductibles, and other limits on actuarially unfair coverage. For example, deductibles optimize the amount of health insurance a person gets by transferring the most catastrophic claims to the insurance company while retaining the lower cost claims (Schlesinger 1981). Other forms of uncertainty that have been used to motivate the optimality of partial insurance include uncertainty over the solvency of the insurer and the payout under the policy. If an individual purchases insurance from an insurance company that may or may not be able to pay claims in full due to solvency concerns, then it would also be optimal for that individual to purchase partial insurance (Munch and Smallwood 1980). These models capture two essential features of health insurance—health insurance is costly to write, and health insurance claims are uncertain. That is, insured individuals who purchase health insurance may not be certain which claims they will incur or whether their insurer is able to pay out claims in the future.

In setting optimal health insurance as a benchmark for health insurance policy, it is also important to consider constraints related to health insurance, Many of the studies in the insurance literature rely on the ability of individuals and insurance companies to choose any level of coverage $0 \leq \alpha \leq 1$ that suits them. In reality,

certain forms of insurance may be considered "unacceptable" by individuals, insurance companies, and health insurance regulators. For example, the optimal insurance policy where an individual must pay all claims below a deductible of $5000 would be difficult or impossible for a person with little or no financial wealth to use. An optimal insurance policy with full or near-full coverage for the individual may be considered as unacceptable by insurers—some insurance companies would refuse to sell such a policy at all for reasons such as "moral hazard" detailed in Sect. 3.3. Finally, regulators may consider a policy that implies premiums that are "unaffordable" to be unacceptable, and will not allow insurers to offer such a policy even if certain consumers were willing to accept such a policy. The economic approach to adding these types of constraints is to look at what is termed "second best" optimization, where a realistic approach to the path dependent, political nature of the health insurance system is in part a determining factor in the type of health insurance trade-offs individuals, organizations, and the governments are able and willing to make.

Researchers have extensively studied the issue of applying the optimal health insurance literature to the problem of determining the optimal scope of health insurance. The optimal health insurance calculations imply a wide range of financial contributions depending in part on the data and assumptions underlying the study. Surveys of the optimal health insurance literature have suggested that the optimal coinsurance ranges from 25 % to as much as 58 % (Cutler and Zeckhauser 2000, p. 587). This α is much higher than the coinsurance rate on many forms of insurance that we will see in this book, including policies as they existed both before and after the ACA. It is important to note that many of these estimates are not "pure" coinsurance policies, since many of the studies also included a "stop-loss" provision. Under a stop-loss provision, coinsurance is limited or eliminated for extremely high cost, catastrophic claims, such as those costing $25,000 or more (Cutler and Zeckhauser 2000, p. 587).

Relatively high coinsurance rates could be seen as either a barrier to optimal insurance or as a benchmark for health insurance policy to move towards. Out-of-pocket expenses of 25 % of healthcare spending would imply a substantial increase in precautionary savings and out-of-pocket payments for much of the U.S. population. However, current savings behavior may be inadequate in order to cover the remaining financial contribution under such a policy. Conversely, out-of-pocket payments as a share of total spending are close to 25 % in many developed countries. "On average, 20 % of health spending is paid directly by patients; this ranges from less than 10 % in the Netherlands and France to over 35 % in Chile, Korea and Mexico" (Organization for Economic Cooperation and Development (OECD) 2013). It is also important to consider that the out-of-pocket share of spending is an average that masks a great deal of variation across the population where some pay a high proportion of their income for out-of-pocket costs, while other spend much less. Variation is especially important in the U.S. where some populations have a high quantity of coverage, while others have a much lower quantity of coverage or no health insurance at all (the "uninsured").

The low take up of low cost or free health insurance also challenges the optimal health insurance literature. The take-up rate of Medicaid, the government subsidized health insurance program for those on low incomes or with certain health conditions, is much less than 100 %. One implication of the economic model presented here is that the U.S. uninsurance rate would be much lower due to a much higher adoption rate for these free or low cost Medicaid policies. Economists have been forced to find explanations for these low take-up rates for insurance that comes at a low cost or is free. "The three main reasons for low take-up, as hypothesized in the literature, are (i) lack of information (information costs); (ii) administrative hassle associated with an application that requires considerable paperwork, verification of income, and visits with caseworkers (process costs); and (iii) stigma associated with public programs (outcome costs) (Craig 1991)" (Aizer 2007). In other words, public health insurance may seem free, but time costs and other burdens associated with non-universal health insurance programs may be a disincentive to signing up for coverage. Alternative solutions such as health insurance mandates or universal health insurance programs have been proposed in the U.S. and implemented in other countries in order to achieve universal coverage. Determining the optimality of mandates or universal health insurance requires the use of techniques to weigh the benefits of policies in terms of increased health insurance against the cost in terms of increased use of health insurance and enforcement of the policies themselves.

3.3 Constraints on the Scope of Insurance

3.3.1 Insurer Constraints

Insurance economics emphasizes the role of risk management as a "supply side" constraint on full insurance. Risk management is important for insurers because of the principles of solvency and financial viability that all insurers must follow. One of the principle goods that health insurers supply is a reserve that is designed to pay for claims that are higher than average (expected). While it may be the case that an insurer may be better able to manage the risk of an individual's healthcare costs due to the insurer's size and financial sophistication, there are limits to this ability. In some sense, the insurer is a "pass through" entity managing the reserve on behalf of the entire insured population. Money paid by an insurer for one individual is not available for another individual, and thus an insurer cannot afford to take on risks that would bankrupt the risk pool as a whole or lead to unacceptably high premiums.

Health services research emphasizes the importance cost-effectiveness as a supply side constraint on full insurance. Cost-effectiveness refers to the amount of health that is obtained for a given level of spending on a particular health service. Cost-effectiveness is a constraint on the supply of health insurance because not all

health shocks can be fully reversed or repaired and because resources to devote to healthcare are limited (Neumann 2004). Given that health insurance pays claims on a service basis rather than paying cash, insurers are put in the position of choosing between what to cover for particular conditions. As a result, the extent of health insurance coverage is limited to an extent by the degree of progress in medical science—the closer that healthcare can bring a person back to their original level of health, the more that insurers are able to provide.

3.3.2 Asymmetry of Information

One of the assumptions included in generic economic models that must be relaxed for health insurance is the assumption of complete information. Complete information means that the buyer and the seller of a good are both fully informed about the characteristics of that good, its price, and its quantity. Complete information is important as a basis of trade. Complete information implies two separate aspects about the understanding of a good. The first is that there is no aspect of the good that is unknown. Complete information is also symmetric—that both buyer and seller have equal amounts of information about the good. It is also possible to for information to be partial but still symmetric, in that both buyer and seller have the same incomplete information about the good being exchanged. Finally, information could be partial and asymmetric—one or both of the parties to a transaction has information that is not shared with the other party (Osborne and Rubinstein 1994).

Health insurance markets feature partial and asymmetric information in a number of ways. Partial information relates to the probabilistic and personal nature of health. Individuals may be well informed about their health risks, but such risks are probabilistic—there is some probability that a given individual will have a heart attack in the next ten years. It is not clear as to whether this risk can be accurately quantified, and it is clear that not all of the information that might be needed to quantify this risk is available. The multifactorial, subjective nature of health capital means that individuals may not be fully informed about their health state at the time that they purchase insurance.

Another important aspect of partial information with respect to health capital is to the different levels of utility individuals might derive from different health states. Two similar individuals who break their ankles might experience very different levels of disutility from the injury. One person might experience a high degree of pain while the other experiences moderate pain and discomfort from the same injury. While economics often relies on these preferences being both stable and discoverable in order to determine the cost of changes in health and the benefits of improvements in health, it might be that no one can truly know what their utility would be having suffered a grievous injury, at least until after suffering the injury. In this sense, defining and measuring "the disutility" associated with future injury or illness would be impossible in many cases, making information about the value of

future health states partial, but not necessarily asymmetric. Uncertainty over the future value of health makes it very difficult to derive preferences over a number of the characteristics of health insurance through consumer choice—consumers simply never have the opportunity to make many of these choices.

There are also strong incentives for both parties to an insurance contract to hold back information in an attempt to take advantage of information asymmetries. Individuals have a strong incentive to disguise their true level of health risk. While misrepresentation of one's health state when signing up for insurance could be fraud, an individual who faces the choice about whether or not to reveal a particular health condition to a health insurer would rationally choose to avoid revealing that information, since it would likely result in a higher premium.[4] Similarly, a health insurer may choose to selectively reveal information about the extent of coverage to individuals considering insurance without excluding required information or misrepresenting itself or violating the standard of utmost good faith described in Chap. 1. Chapter 7 further describes how health insurance includes a number of characteristics that consumers do not consider, or would not have the ability to consider ahead of time, because health insurance is often purchased by an organization on behalf of a large group.

Another result of information asymmetries is the possibility that insurance providers may behave in a manner that interferes with market function. Health insurance markets feature a range of different insurers, some with greater quality and some with lesser quality. If that quality is difficult to distinguish *ex ante* by a consumer, then higher quality insurers will have trouble commanding a higher price, and the overall quality of insurance may suffer (Armstrong and Chen 2009). This difficulty is heightened in health insurance both because of the subjective nature of healthcare claims and because many of the providers of health insurance are not traditional health insurance companies, but rather employers, governments, or healthcare providers.

Asymmetric information is also problematic to the role of healthcare providers as experts in the provision of medical care. Health economics recognizes the general problem that providers could profit from an informational advantage ("supplier induced demand" or SID) when providing healthcare services to an individual. The third-party payment system for health insurance whereby payments for care from an insurer rather than directly from the patient might give providers an additional channel from which to benefit from their informational advantage, since providers can profit from the additional care provided without imposing a direct financial burden on their patients (Pauly and Satterthwaite 1981).

[4]Of course, insurers can ask consumers questions, and then rate the insurance on the basis of individual characteristics (underwriting). There is always a limit to the amount of information a company can obtain from a consumer and use, as well as the possibility that the consumer would misrepresent themselves, i.e., commit fraud. One of the major changes under the ACA is the end of such underwriting in the nongroup market. However, such underwriting has not been a feature of the group or government-provided markets for years (or ever).

3.3.3 Moral Hazard

Moral hazard is a phenomenon related to the effect of insurance on the price of healthcare services. Economics emphasizes the role of prices in helping individuals to make consumption decisions in the present, and to make the intertemporal substitution decision about what to consume today and what to save for tomorrow. Under certain assumptions, allowing consumers and suppliers to freely trade for goods will lead to a market price that leads to an efficient allocation of economic resources. However, one of the assumptions underlying the benefits to free trade is rationality, which gives rise to moral hazard. Pauly (1968) emphasizes the rational basis of "moral hazard": "… that the response of seeking more medical care with insurance than in its absence is a result not of moral perfidy, but of rational economic behavior." Second, he points out the role of individual preferences in determining the optimal health insurance contract, or rather contracts, since individual preferences including the degree of risk aversion determine the optimal contract.

The main concern with moral hazard as a price effect is that health insurance may induce inefficient care or coverage. After an individual has faced a loss, any form of insurance that pays for healthcare services directly decreases the price of healthcare. One popular analogy is to a restaurant meal—if a group of five agrees to split the bill evenly regardless of what they purchase, then there is a tendency to "run up" the bill (i.e., spend more freely) since each person's additional $1 in spending only adds $0.20 to their own personal share of the check. In insurance in general, the concern is that the marginal cost of covered losses will exceed the marginal benefits. If health insurance changes the price of healthcare by lowering it, the effect could be to raise the consumption of healthcare to a high (inefficient) level. "Over consuming" healthcare because an insurer is paying part, or all, of the bill, could be seen as a cost that drives everyone's premiums up.

Moral hazard is also a particularly important problem in health insurance for several reasons. Healthcare is often reimbursed on a service basis rather than being paid in cash to the insured individual, which has the effect of shielding individuals from the cash price of the care they receive. Health capital is also a form of capital whose value has no natural limit. In contrast, property and casualty insurers have a natural limit to the overall size of losses—an automobile insurer can declare a "total loss" for an automobile and cap the payout at the market value (or replacement value) of the automobile. Health insurers have a more difficult time employing such a strategy because the value of human life is not a market commodity.

Moral hazard has a direct economic cost in terms of the scope of insurance. The direct economic losses related to moral hazard include the cost of providing insurance coverage and the value of care that is paid for by health insurance. Claims costs for insured individuals are higher than the cost of providing the equivalent amount healthcare individuals without using health insurance because of the cost of

utilizing the insurance claims adjudication system.[5] This means that the premiums for health insurance are higher than the average cost per person for self-paid healthcare spending. This higher spending has the effect of increasing the amount of an individual's budget that is devoted to healthcare when they obtain insurance, thus reducing the amount available for other forms of consumption. However, this does not necessarily reduce the value of insurance, since individuals who pay a higher premium and receive more care are, in some sense, getting what they pay for. However, to the extent that loading costs are related to the overall amount of claims, moral hazard that inflates the size of payouts will also impose a higher cost of providing insurance on individuals. Variable costs of health insurance that are increased by moral hazard represent a real cost to individuals and to other purchasers of insurance.

Moral hazard also has a direct economic cost and an indirect cost in terms of the scope of insurance. Moral hazard reduces the value of care in the sense that the marginal additional care that a person receives through health insurance may have low marginal benefit. A person who is personally paying for healthcare would tend to purchase the healthcare with the highest marginal benefit. Any program that increases an individual's budget for healthcare spending could result in their purchasing care with a lower marginal benefit. This is the sense in which "inefficient care" generally refers to care that an individual obtains because of the presence of health insurance but would not have chosen to obtain in the absence of health insurance.

The existence of moral hazard is one of the main rationales given for the absence of full insurance, or full coverage of losses (Hölmstrom 1979; Winter 2000). Given that consumers faced with insurance will generally consume more than they would in the absence of insurance, health insurers will price their insurance to include both individual preferences for healthcare in the absence of insurance and the price effect associated with moral hazard. Consumers, in turn, will factor this higher price into their decision-making about the purchase of insurance. It is difficult but not impossible to measure the impact of moral hazard on the scope of health insurance available in health insurance markets (Chiappori and Salanie 2000).

The desire to focus health insurance spending on the highest value services is one of the rationales for third-party payment, physician licensing, and partial insurance. Third-party payment systems are often viewed as a way to reduce moral hazard by taking the control over payment determination away from individuals or physicians (Arrow 1963). Alternatively, licensing and certification of physicians and healthcare professionals can limit the moral hazard problem if the licensing system includes mechanisms to discourage lower value care. Using a small copayment can raise the price of care, thus making the price that an individual with

[5]Note that this assumes that all else is equal between the two claimants, which is clearly not the case. For example, an insurer could use its scale to negotiate a lower price for a given service, thereby paying less when adding together the lower negotiated rate and the cost of adjudicating the claim. That is why the economies of scale and scope are crucial to justifying the use of health insurance on an economic basis.

insurance faces closer to the full price of healthcare and reducing the price effect that encourages those with health insurance to obtain additional healthcare.

3.3.4 Adverse Selection

Adverse selection refers to the tendency for those individuals who are more likely to make an insurance claim to seek insurance in the first place. Adverse selection [also known in the actuarial literature as "antiselection" (Bluhm 2007, Chap. 4)] can be viewed as a rational response to private information as with moral hazard, given that it is in people's best interest not to reveal information that will lead to higher premiums, and that we should expect that individuals will not volunteer such information unless asked.[6] Adverse selection is an important consideration in insurance because it imposes costs on others: when individuals pay a premium less than their expected costs, then other members of the insurance pool are left to make up the difference.

Adverse selection is distinguished from moral hazard in that adverse selection occurs before a person obtains insurance, whereas moral hazard occurs after a person has insurance.[7] Bluhm (2007), in his discussion of "managing antiselection" distinguishes between "external," "internal", and "durational" antiselection. These distinguish between the tendency of a person to initially enter an insurance policy with knowledge that they are likely to incur a claim ("external" antiselection), the behavior of individuals in response to insurer offers regarding renewal of insurance ("internal" antiselection), and the tendency of those with higher claims to be more likely to retain their insurance over time ("cumulative" antiselection) (Bluhm 2007, pp. 84–85).

Several economic models apply to the study of how adverse selection could limit the scope of health insurance. The main concern with adverse selection is that health insurers may undersupply insurance in response to consumer use of private information. Recall that insurance is provided in a market and that producers not only respond to consumers, but rationally attempt to anticipate their behavior and factor it into their offers of insurance. Two classical models set the foundation for the examination of adverse selection in general.

One important adverse selection model is the Akerlof "lemons" model (Akerlof 1970). In it, Akerlof considers the market for used cars and the possibility of asymmetric information (sellers know more about cars than buyers). He demonstrates that none of the cars may be sold despite the fact that all the sellers are willing to accept less than the buyers are willing to offer due to the fact that buyers

[6]Misrepresenting oneself when the health insurer does ask about health status or other variables is a form of fraud as discussed above concerning underwriting (note iv).

[7]It is also important to note that this is not a black and white distinction. "Ex ante moral hazard" refers to a situation where a person does not undertake preventative behavior in anticipation of being able to address the need for care through insurance in the future (Ehrlich and Becker 1972).

cannot distinguish between the high and low quality cars (the lemons). Applying this to health insurance, the model implies that health insurance might not be available in a free market despite the fact that consumers are willing to pay more than the minimum insurers demand. Akerlof considers this possibility to be an argument in favor of Medicare, whereby the government offers insurance to older individuals, some of whom might be considered as "uninsurable" by health insurers.

Rothschild and Stiglitz (1976), consider the situation in a competitive insurance market which features private information. Individuals are free to accept or reject any insurance contract, and they are motivated to purchase insurance due to risk aversion. Insurers are also free to offer any insurance contract, and their main constraint is that a competitive market "… involves free entry and noncollusive behavior among the participants in the market". The result is a "separating equilibrium" where those at higher risk purchase an insurance contract with relatively more protection against risks at a higher premium, whereas those at lower risk purchase an insurance contract with relatively less protection against risk at a lower premium. Crucially, insurance companies make a zero profit in this model—the separation into two classes of insurance is the best the market can do, but it does not result in positive economic profits (rents) for the insurance company. This model is also important in the manner in which insurers induce individuals to "reveal" their expected claims. Offering different levels of coverage to different individuals at different premium levels is one way for the insurers to determine who is likely to be a high-risk or low-risk individual even if individuals do not wish to share this information.

Empirical studies of adverse selection have shown how this rational tendency can substantially change the scope of insurance. A striking example of an "adverse selection death spiral" is given by Cutler and Reber (1998). It is called a "death spiral" because adverse selection may make certain health insurance contracts unsustainable, leading to insurers withdrawing these insurance offerings. This study Investigated the impact of this change in Harvard University's benefit structure, in which the university moved from making variable contributions to employer-provided health insurance options to making a fixed contribution (also known as a "voucher" system). Employees quickly switched from the higher cost "PPO" plan to the lower cost HMO plan, and this switch started with those employees who were younger and, presumably, healthier. Harvard was forced to drop the PPO plan after 3 years. The authors concluded that the economic result was a welfare loss to enrollees due to adverse selection, combined with a transfer from insurance companies to Harvard due to lower premiums for the less costly plans that remained.

A recent review by Cohen and Siegelman (2010) notes both a number of prior studies that have tested for adverse selection in health insurance. These prior studies have found positive results in some cases (the presence of adverse selection) and negative results in others (studies that tested for and did not find adverse selection). They go on to explore reasons that the theory of adverse selection may not be borne out in practice—individuals may not be well-informed about their health status or

how it translates into insured benefits, and insurance companies may have an informational advantage due to data and sophisticated predictive techniques. The authors also raise the possibility of "propitious selection" whereby those who are more risk averse, and thus more likely to purchase insurance, also have a lower risk related to their health. This relates to an issue, seen throughout studies of insurance, that many results may be a result of unobserved variation in risk aversion, unobservable differences in health status, and diversity in preferences over various states of health, leading to inconclusive results with respect to adverse selection.

In part, the mixed results found in studies of adverse selection and moral hazards are due to the difficulty in disentangling these two effects in practice. For example, Rothschild and Stiglitz (1976) introduce their study, in part, by raising concerns about moral hazard—the term "adverse selection" does not appear in their paper (instead, they focus on the separating equilibrium as a type of "self-selection" model) (Rothschild and Stiglitz 1976). Empirical work that relies on the choices of individuals in terms of their plan is not able to control for unobservable differences among the population in an insurance plan. Given that moral hazard and adverse selection are issues of information asymmetry, the individuals in these studies are presumably concealing their underlying preferences from the researcher as well as the health insurer.

3.3.5 Other Economic Externalities

In economics, "externalities" refers to the possibility that a consumer may not pay the full price of a service or that a producer may not pay the full cost of producing a services. This is a violation of an assumption that is embedded into the unconstrained optimization approach, where consumers and producers "internalize" (i.e., feel the effects of) all the consequences of their choices. The main consequence of a consumer's choice is the payment of the price, which means the money or traded resource, cannot be used for another purpose. The main consequence of a producer's choice is the cost of providing the good or service provided—the producer should carry or account for the entire cost of producing the good or service. In health insurance, this would mean that consumers would be expected to pay the full cost of their health insurance, and that producers would be fully responsible for the cost of providing that insurance.

Two of the most important externalities in health insurance are moral hazard and adverse selection. Moral hazard and adverse selection impose a welfare loss on other consumers or producers to the extent that they experience some of the consequences of the individual's behavior. An individual who consumes an inefficient amount of healthcare in some sense imposes that cost on the others in the group—society experiences a loss because the money allocated to the low value service would have been better spent on something else. Adverse selection could be seen as imposing a welfare cost on other consumers or on producers. Individuals who use their informational advantage to buy health insurance at an actuarially favorable

premium must have their care subsidized either by other members of the plan or through the health insurer suffering a loss.

The ability of people to obtain care that they are unable to pay for can also be considered an economic externality. For example, there is often an expectation, and often a legal requirement, that providers and facilities will treat patients without regard to their ability to pay. Special expectations of physicians in terms of their behavior is one of the unique characteristics of the healthcare market as described in Arrow (1963), in his examination of the economics of the healthcare system. Hospitals are bound by laws such as EMTALA (the Emergency Medical Treatment and Active Labor Act) to provide care to those suffering "… acute symptoms of sufficient severity (including severe pain) …" as well as "… to a pregnant women [woman] who is having contractions" (Lee 2004). In the absence of such guarantees, individuals would rationally anticipate the cost of care, and plan for them through the use of health insurance or precautionary savings. One concern about these guarantees is that the incidence of these healthcare costs for those who cannot afford to pay falls in part on those who do have health insurance, an externality. To the extent that health insurance pays for the costs of those without the means to pay, the uninsured impose an externality cost on those with health insurance. This cost is one of the major justifications for health insurance mandates. However, the magnitude and importance of the costs of the uninsured for those with health insurance is one of the most important debates in the health economics and health services literatures (Gruber 2008).

Incomplete markets are another externality that may prevent individuals from obtaining the insurance they want. Incomplete markets are a general concern in the insurance literature relating to the fundamental assumptions underlying financial models. Many of these models assume that any risky contingency can be bought or sold as a "contingent claim" or financial security (Huang and Litzenberger 1988). In the context of health insurance, complete markets would imply that a market exists for any possible risk related to health. Complete markets are important because they support the insurance choice whereby an individual can purchase any level of protection or an insurance company can sell any level of protection α. In the absence of complete markets, certain levels of protection may not be available at any price. Incomplete securities markets could be considered a market failure that imposes costs on the individual or insurer who are unable to write a mutually agreeable contract due to the market's failure to reinsure part of the risk they are taking on, leading to individuals receiving less insurance than they would like.

In a related issue, health insurance itself is risky in that the health insurer may fail to pay benefits or claims as promised. The inability of an insurer to pay claims can be considered an externality because the failure of the insurer may be passed on to other parties, such as governments that take on the costs of bankrupt insurers. In this case, the individual is less "insured" than he or she would like to be, since reducing their risk through health insurance involves simultaneously taking on a smaller but nonzero risk that the insurer will fail. This market failure imposes costs on the insured individual or providers counting on the insurer to pay certain claims, as well as imposing costs on taxpayers who foot the bill for bankrupt insurers.

The actual nonpayment of claims in health insurance is often less of a concern than general issues of viability that include the timeliness of claims payments to providers or strictness of insurer review of claims.

Many of the financial aspects of health insurance arose or are specifically designed to ensure that participants in health insurance markets internalize the costs of their behavior. Internalizing the cost of informational advantages that individuals have with respect to their own health status is one of the main justifications for the use of partial insurance. The use of financial responsibility for individuals with insurance could be seen as a way to get them to pay partially for the cost of care that results from personal risk or behavior, often referred to as having "skin in the game" (e.g., Neuman et al. 2007). However, the use of financial responsibility as a response to asymmetric information is imperfect. Financial responsibility based on the dollar cost of healthcare does not distinguish between more and less effective care. There is also an important issue related to the limits to what lower income individuals can pay out of pocket. When individuals cannot or will not pay those costs, the costs are instead born by society at large, thereby externalizing costs.

Similarly, the history of getting producers of healthcare to internalize the cost of health insurance has been marked by both successes and failures. One example of a system that is designed to place the responsibility for the utilization of care on healthcare providers is the prospective payment system (PPS). The PPS was devised by the Medicare program as a way to have providers internalize the costs of the care that they provided by paying hospitals a fixed rate for episodes of care. The system has been measured in terms of reduced length of stay (LOS), which would indicate a reduction in part of the inefficient overutilization of care induced by health insurance. Prior studies have found mixed results on the effectiveness of the PPS, indicating that the effect on LOS was moderate or perhaps zero (Epstein et al. 1991; Newhouse and Byrne 1988). The main cost of the PPS and related systems is that it imposes a higher loading cost on insurance since it is a more complex system. There is also substantial evidence that providers can continue to use their informational advantage to provide both of inefficiently high levels of treatment and inefficiently low levels of treatment for certain groups and conditions in order to maximize the profits they obtain from Medicare and other insurers, sometimes referred to as "upcoding" (Silverman and Skinner 2004). In that case, society continues to pay for the externalities arising from third-party payment. Ultimately, the incidence of such costs and which system minimizes costs to society are crucial determinants of the optimal health insurance system and who pays for it.[8]

[8]There is also a more political question in terms of who derives the most benefit from the current system as it is structured. That question is important because it influences and constrains possible solutions to improve health insurance and health insurance markets. For an in-depth examination, Steven Brill's "America's Bitter Pill: Money, Politics, Backroom Deals, and the Fight to Fix Our Broken Healthcare System" serves as an account of how many of these forces shaped the ACA (Brill 2015).

3.3.6 *Irrationality of Consumers and Producers*

There is one final issue that may limit the scope of health insurance: that many individuals are not fully rational. Not all economic decisions are made on a purely forward-looking basis, violating one of the main assumptions underlying expected utility maximization. Important behavioral biases for insurance include loss aversion, framing effects, and zero-risk bias. Loss aversion, the tendency to experience more disutility losses than the utility experienced from gains, may cause consumers to overvalue health insurance. Framing effects, which is the tendency to judge risks inconsistently based on how they are presented, can lead individuals to get different amounts of utility from two equivalent health insurance contracts if terms are explained in different ways. Zero-risk bias, the tendency to see a small risk as a non-risk, may lead individuals to undervalue the extent to which health insurance can protect them from catastrophic events (Kahneman 2003). Each of these biases undermines the result that risk averse individuals will eagerly purchase actuarially fair insurance. These limitations of classical economic models in insurance are often grouped under the term "behavioral economics" or "behavioral insurance" (Kunreuther et al. 2013; Richter et al. 2014).

Empirical studies have shown the power of behavioral economics to explain health insurance choices that are not explained by classical economics. For example, status quo bias, the tendency to prefer things to remain as they are, results in individuals keeping the health insurance they have rather than switching to a better policy. Handel (2013) demonstrated status quo bias with a case in which an employer offered several policies to employees, including one that policy was clearly inferior to (dominated by) another. The psychic cost of switching can be seen in the number of individuals that remained in the inferior plan, despite the fact that switching did not result in a financial cost to them. The alternative, behavioral economic explanation is that those in the plan that became inferior developed some "attachment" or had a higher value for the plan as a result of having previously been enrolled in it. This attachment may limit the scope of health insurance by reducing the perceived value of new health insurance policies relative to existing policies.

Another form of consumer demand that is at odds with the rational framework iOS the popularity of coverage for low severity, high frequency events. The models of insurance demand imply that insurance should generally be both partial and concentrated in terms of coverage for catastrophic events. The reason that this is true generally is that full insurance may be optimal when markets are complete, insurance is risk-free, and information is perfectly symmetric. Partial insurance for catastrophic events would pay for high severity, low frequency costs such as suffering a heart attack and exclude low severity or high frequency claims. Such insurance is generally quite unpopular—for example, one of the main complaints with Medicare Part D drug coverage was the "donut hole" or limited coverage for spending on smaller drug bills. This motivated a change to Medicare Part D under the ACA by "closing" this donut hole (Centers for Medicare and Medicaid Services 2015). Again, this limits the scope of insurance since it discourages insurers from offering particular types of partial health insurance to individuals.

This concludes the section focused on the importance of health insurance. The next section will focus on health insurance markets, meaning an examination of how health insurance is actually set up. It explores the economics of demand for insurance, and supply of insurance, and the institutional organization of the health insurance industry. The rationale for examining health insurance in this way flows explicitly from the recognition that what is optimal in health insurance is "second best health insurance." We take the behavior and preferences of individuals, organizations, and the government as given rather than attempting creates a system that would be optimal in a vacuum. Ultimately, the description of health insurance markets in Sect. 3.2 motivate the examination of health insurance policy in Sect. 3.3. Improvements in health insurance policy involve making changes that can move the health insurance system towards the optimal benchmark described in this chapter. In economics, this is also known as an exercise in "positive" rather than "normative" economics. This means a description of the economics of what could be rather than the economics of what should be.[9] The next step is a view of the demand for health insurance in Chap. 4.

References

Aizer, A. (2007). Public health insurance, program take-up, and child health. *The Review of Economics and Statistics, 89*(3), 400–415.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics, 84*(3), 488–500.

Armstrong, M., & Chen, Y. (2009). Inattentive consumers and product quality. *Journal of the European Economic Association, 7*(2–3), 411–422.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review, 53*(5), 941–973.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., et al. (2013). The oregon experiment—effects of medicaid on clinical outcomes. *New England Journal of Medicine, 368*(18), 1713–1722.

Bluhm, W. F. (2007). *Individual health insurance*. Winsted, CT: ACTEX Publications.

Brill, S. (2015). *America's bitter pill: Money, politics, backroom deals, and the fight to fix our broken healthcare system*. New York: Random House.

Centers for Medicare and Medicaid Services. (2015). *Closing the coverage Gap—Medicare prescription drugs are becoming more affordable*. (No. 11493). Washington, D.C.: Department of Health and Human Services.

Centers for Medicare and Medicaid Services. (2016). *Information on essential health benefits (EHB) benchmark plans*. Retrieved from https://www.cms.gov/cciio/resources/data-resources/ehb.html

Chernew, M. E., Rosen, A. B., & Fendrick, A. M. (2007). Value-based insurance design. *Health Affairs, 26*(2), w195–w203.

[9]Milton Friedman's "Essays on Positive Economics" providers one extended treatment of positive versus normative economics, with an emphasis on the use of the positive economics method (Friedman 1953).