# Chapter 18
# Health Applications

Anikó Bíró, Mark N. Harris, Andrew M. Jones and Nigel Rice

**Abstract** The chapter outlines the history of health econometrics from the pioneering development of new empirical techniques for the RAND Health Insurance Experiment, through the adoption of microeconometric approaches driven by the availability of data from large-scale surveys, household panels and linked administrative datasets, on to the influence of the credibility revolution and a focus on quasi-experimental approaches to impact evaluation. Overall, developments in the use of econometric methods within health economics have reflected the unique features of health data and the aspiration to inform evidence-based policy and decision making.

## 18.1 Introduction

This chapter provides an overview of major developments in the application of econometric methods in health economics. Broadly speaking, health economics is concerned with the efficiency, effectiveness, value, and behaviour involved in the production and consumption of health and healthcare services. Common themes span the design of efficient and fair healthcare systems, incentives in payment mechanisms and resource allocation, the cost-effective use of limited healthcare resources, and the evaluation of healthcare and public health interventions. Key aspects and challenges of its study include information deficits and asymmetries, uncertainty in health risks,

Anikó Bíró ✉
ELTE Centre for Economic and Regional Studies, Budapest, Hungary, e-mail: biro.aniko@krtk.elte.hu

Mark N. Harris
Curtin University, Perth, Australia, e-mail: Mark.Harris@curtin.edu.au

Andrew M. Jones
University of York, York, UK, e-mail: andrew.jones@york.ac.uk

Nigel Rice
University of York, York, UK, e-mail: nigel.rice@york.ac.uk

third party payers, and externalities. An increased emphasis on evidence-based policy making over the past decades, coupled with increased access to sources of rich administrative and survey data has fuelled the need for sophisticated and considered empirical investigation.

The label *health econometrics* has been adopted as a convenient short-hand to describe the development and application of econometric methods within health economics. The label was used as the title for a handbook chapter by Jones (2000), although earlier authors had reviewed the use of econometrics in the field (see Newhouse, 1987; Wagstaff, 1989). An early exponent of econometric methods with health data was Feldstein (1967), who used data on British hospital costs. Another pioneering paper was Auster, Leveson and Sarachek's (1969) analysis of death rates across the United States in 1960 that used instrumental variables (IV). The RAND Health Insurance Experiment (HIE) was the catalyst for much of the methodological innovation that took place in the 1970s and 1980s (Manning, Newhouse, Duan, Keeler & Leibowitz, 1987). The scale of activity grew substantially over subsequent decades. The *European Workshops on Econometrics and Health Economics*, established in 1992, provided a focus for these developments and networking by researchers in the field. These workshops have been complemented by similar meetings in North America, Australasia and Asia.

The challenges posed by health data have stimulated important methodological innovations. There has been a dramatic growth in econometric studies that use health data. This has stimulated developments in econometric methodology that have spread beyond health economics. Health econometrics has developed and applied empirical methods that are tailored to the analysis of health and healthcare data. Distinct and recurring statistical features of outcomes and policy variables of interest provide a particularly rich field for the application of econometric techniques. These include issues of latent variables, unobserved heterogeneity, endogeneity, nonlinear models, and methods for handling survey and item non-response.

As in other sub-fields of economics, econometric methods used in health economics, while tailored to the specific challenges outlined above, have often been adapted from other fields within economics, as well as from the broader social sciences and statistics. In this context, health econometrics cannot be considered in isolation as having developed a bespoke set of methods. However, the unique features of health and healthcare, that make health economics an important and valuable sub-discipline of economics, have lent themselves to a focus on particular techniques. Many of the developments in techniques have arisen from the nuances of new directions of research and the specific outcomes and statistical challenges they entail, as well as advances in data availability and their richness. For example, the increased use of panel data from the 1990s and of administrative sources of data from the 2000s onwards.

A strict chronological ordering of the application and developments of econometric methods in health economics would appear superfluous. Instead we categorise developments into broad eras, or epochs, in which methods first became prominent in the applied literature. Further extensions of these approaches are then described. Many of the techniques outlined remain relevant today and while some methods

have experienced intense periods of interest, book-ended by hiatuses, few have faded completely from the analytical toolkit.

The chapter commences by describing a pivotal period in the development of health econometrics – the RAND Health Insurance Experiment (HIE) in the mid-1970s – which represented a formative moment in recognising the need for empirical approaches tailored to the many peculiarities that health and healthcare data present. In particular, this fostered an emphasis on limited dependent variable models that were designed to account for decision-making when accessing care, and the related problem of modelling the skewed distribution of healthcare expenditures. This is followed, in Section 18.3, by examining approaches that were used to deal with qualitative outcomes, predominantly binary, ordered categorical and interval data, and extended to consider multinomial choice models typically used for analysing decisions, for example, over insurance plans or healthcare providers. The development and application of count data models for health service interactions, and duration models used to capture both occurrence and timing of health-related events are described; noting that health data have provided a catalyst for the development of these methods in econometrics. Section 18.4 turns to approaches that gained prominence with increased access to richer information available from social surveys and panel datasets. These include methods to address self-reporting bias inherent in many individual measures of health status, and to model the inherent unobservable latent nature of health; developments in the measurement of health and healthcare inequality; and approaches to spatial spillovers in the delivery and organisation of health services. The ability to account for individual heterogeneity and dynamic relationships afforded by increasingly long panels is also described together with extensions to nonlinear models. Section 18.5 describes approaches to causal inference which, in common with other applied areas of economics, has become a dominant focus of recent empirical research. This focuses on methods for observational data and the growing emphasis on quasi-experimental study designs. The chapter finishes with a discussion (Section 18.6) including neglected topics and prospects for future research. Section 18.7 concludes.

## 18.2 The RAND Experiment and Its Influence (1970s–1980s)

The RAND HIE remains one of the most influential studies in health economics and public policy. Conducted between 1974 and 1981, the experiment randomly assigned more than 5,800 individuals from roughly 2,000 households to health insurance plans with varying levels of cost sharing. This randomized design allowed researchers to credibly estimate the causal impact of insurance coverage on healthcare utilization and spending, overcoming the selection problems that typically complicate observational studies of health insurance markets (Manning et al., 1987). As later emphasized by Aron-Dine, Einav and Finkelstein (2013), this represented a pioneering application of randomized experimental methods in the social sciences and in the economic analysis of moral hazard in health insurance.

The primary legacy of the experiment is its evidence on how the price of medical care affects utilization. The experiment demonstrated that individuals facing lower out-of-pocket costs consume more healthcare services, with an estimated price elasticity of demand for medical spending of roughly $-0.2$. These results provided one of the first credible quantitative estimates of the extent of moral hazard in health insurance and have since played a central role in both academic research and policy analysis. Indeed, even decades later, the results are often viewed as the 'gold standard' evidence for predicting how changes in insurance generosity will affect medical spending and for guiding the design of insurance contracts (Aron-Dine et al., 2013).

Beyond its substantive findings, the HIE has had a lasting influence on empirical work in health economics. The experiment produced an unusually rich dataset on health outcomes, healthcare utilization, spending, and insurance design, which has enabled extensive reanalysis and methodological developments. Moreover, the study helped establish randomized experiments as a powerful tool for evaluating social policies. Since large-scale experiments are extremely costly and difficult to implement, it remains a uniquely important source of evidence on the demand-side behavioural response to health insurance coverage.

### 18.2.1 Models for Limited Dependent Variables (1980s-)

A key empirical problem addressed by the RAND HIE involved the consideration of *limited dependent variables*, where the outcome of interest contained a large proportion of zeros and then (roughly) continuous positive values. Examples in health economics go back to Rosett and Huang (1973) and persist to this day (Enami & Mullahy, 2009). Limited dependent variables include medical expenditures, physician visits, or health service utilization.

At the time, three major modelling strategies emerged to address limited dependent variables: the *two-part model*, *sample selection (selectivity) models*, and *hurdle models*. These approaches differ in their assumptions about the data-generating process, particularly regarding the interpretation of zero outcomes and the relationship between participation and intensity decisions. Where participation and intensity are considered to be determined sequentially and independently, two-part models (sometimes also referred to as hurdle-models) are appropriate. Sample selection models are applied where zero (or more correctly, *unobserved*) outcomes arise due to the continuous variable not being observed when 'participation' does not occur. Where participation and intensity decisions are jointly determined but must both be satisfied for positive outcomes, double hurdle-models are preferred. This basic taxonomy highlights that the choice of model depends on the underlying behavioural process rather than purely statistical considerations.

#### 18.2.1.1 Two-Part Models

The influential contribution of the two-part model (2PM) reflected an important conceptual shift in how economists understood individual decision-making in the context of healthcare demand. Rather than viewing utilization as a single, continuous choice, the two-part model recognized that individuals first decide whether to seek care at all, and only subsequently determine the level of consumption.

This framework emerged as researchers tried to address the unusual features of health expenditure data, especially the many zero values and the highly skewed distribution of positive expenditures. Traditional econometric models, such as the Tobit or sample selection models, imposed relatively strong assumptions about the underlying data-generating process. In contrast, the two-part model offered a more flexible alternative by allowing the participation decision and the intensity of use to be modelled separately. This separation meant that different econometric techniques and functional forms could be applied to each stage, reflecting potentially different behavioural mechanisms.

The prominence of the two-part model in health econometrics can be traced back to its application in the analysis of the RAND HIE (Newhouse, Phelps & Marquis, 1980; Manning et al., 1987). Building on this work, Duan, Manning, Morris and Newhouse (1983) strongly advocated for the approach, emphasizing its relatively weak distributional assumptions and practical advantages in applied research. The model quickly became a standard tool for analysing healthcare expenditures, particularly in policy contexts where accurate predictions of mean spending were essential.

A key feature of the two-part model is that it decomposes expected healthcare expenditure into two independent components: the probability of any use, and the expected level of expenditure conditional on use. This decomposition proved especially valuable for policy analysis, as it allowed researchers to examine how different factors influence access to care separately from the intensity of utilization (Pohlmeier & Ulrich, 1995). Such distinctions are often central to evaluating health policies aimed at improving access or controlling costs.

As the methodology evolved, attention turned to practical issues arising from specific modelling choices. A common empirical strategy involved modelling positive expenditures using a logarithmic transformation, which addressed skewness in the data. However, this introduced the problem of retransformation bias when converting predictions back to the original scale. To address this, Duan et al. (1983) proposed the smearing estimator, a nonparametric correction that enabled consistent estimation without relying on restrictive assumptions such as log-normality. Subsequent research further refined these methods. For example, Manning (1998) demonstrated that heteroscedasticity could compromise the consistency of the smearing estimator, prompting the development of more robust retransformation techniques (also, see Mullahy, 1998). These contributions highlight a broader theme in the history of health econometrics: the continuous interplay between empirical challenges and methodological innovation.

Importantly, the flexibility of the two-part model has remained one of its defining strengths. While early applications often relied on log-linear specifications in the

second stage, the framework itself places few restrictions on the choice of model for positive expenditures. Any appropriate specification defined over the positive real line can be employed, allowing researchers to tailor their approach to the specific characteristics of the data. We return to such approaches in Section 18.2.2. This adaptability has ensured the enduring relevance of the two-part model in health economics and its continued use in empirical research.

#### 18.2.1.2 Sample Selection Models

The use of econometric methods in health economics has also been shaped by situations in which outcomes are only observed for a subset of individuals. This is a common feature of health data, where measures such as expenditures, utilization, or insurance choices are only defined for those who engage with the healthcare system. To address this issue, researchers turned to sample selection models, most notably the framework developed by Heckman (1979). The approach recognized that the process determining whether an outcome is observed is not random. Individuals first make a participation decision, such as whether to seek care or purchase insurance, and only if this occurs is the corresponding outcome realized. The key insight of the Heckman model was that unobserved factors influencing participation may also affect the level of the outcome itself. In the context of health economics, these unobserved factors might include underlying health status, preferences for risk, or access to information.

Sample selection models were quickly adopted in early health economics applications. For instance, Van de Ven and Van Praag (1981) used this framework to study preferences for health insurance deductibles, highlighting how selection into different insurance plans could bias observed choices. Subsequent research extended these methods to analyse physician visits, hospital utilization, and other forms of healthcare demand, where the decision to seek care and the intensity of use are often closely intertwined. Their development marked an important step in the broader effort within health econometrics to deal with non-random samples and to understand better the mechanisms shaping healthcare decisions.

A major methodological debate in health economics ensued concerning whether two-part or sample selection models provide a better representation of healthcare demand. Duan et al. (1983) argued that two-part models are preferable because sample selection models rely on strong distributional assumptions and can suffer from numerical instability. They also demonstrated that, for many policy questions, both approaches produce similar predictions of mean expenditures. In contrast, Hay and Olsen (1984) and Maddala (1985) argued that the two-part model imposes unrealistic independence assumptions between participation and intensity decisions. They suggested that when these decisions are jointly determined, sample selection models provide a more appropriate framework. Monte Carlo studies by Manning et al. (1987) initially supported the robustness of the two-part model for predicting conditional means. However, later work by Leung and Yu (1996) showed that the apparent superiority of the two-part model may arise from collinearity problems in sample selection models rather than fundamental differences in model performance.

Overall, the consensus that emerged from this literature wss that model choice should depend on the economic structure of the decision process rather than purely statistical considerations.

#### 18.2.1.3 Hurdle Models

Another development in the econometric analysis of health-related outcomes was the class of models commonly referred to as hurdle models (Cragg, 1971; Deaton & Irish, 1984). While the term 'hurdle model' is widely used, the label 'double-hurdle' more accurately captures the underlying idea: individuals must overcome two separate barriers before a positive outcome is observed. This contrasts with the two-part model, which can be thought of as involving only a single hurdle.

Hurdle models were also motivated by the need to address health data characterized by a large number of zero observations, while offering a different interpretation of zero outcomes. The central feature of the approach is that a positive outcome is only observed if two separate conditions are met. The first hurdle represents the decision to participate, while the second governs the level or intensity of the outcome. If either of these conditions is not satisfied, the observed outcome remains zero. This structure implies that zero observations contain information about both stages of the decision-making process, rather than being associated solely with non-participation. By allowing zeros to inform both stages, hurdle models can capture more complex behavioural patterns, particularly in settings where the absence of activity may arise from multiple underlying mechanisms.

Hurdle models have proved particularly influential in the analysis of count data. Applications included modelling the number of physician visits or other discrete measures of healthcare use, where excess zeros are a common feature (Gerdtham, 1997; Gurmu, 1997). One of the earliest applications to health-related behaviour is found in Jones (1989), who employed this framework to study smoking behaviour, illustrating its usefulness beyond traditional utilization measures.

A number of developments that extended the above modelling approaches took place in the 1990s and beyond. First, semiparametric estimators that relax distributional assumptions in selection models were considered (Stern, 1996; L.-F. Lee, Rosenzweig & Pitt, 1997). Second, through methods based on covariance restrictions and partial identification that provide alternative ways to address selection bias without relying on strong parametric assumptions (Pitt, 1997). A further development involved bounding approaches, such as those proposed by Manski (1993), which provide estimates of treatment effects without full identification. These innovations reflected a broader trend toward methods that were more robust to functional form and distributional assumptions. In general though, both double-hurdle and sample selection techniques have fallen-out of favour with applied researchers in the broad field of health economics. The primary reason appears to be that, whilst being based on solid theoretical grounds, and not, in general, being explicitly reliant on exclusion restrictions, identification is much tighter when these are available. However, in practice, it remains challenging to find convincing identifying variables for either part(s) of the model. However,

two-part models have remained a strong workhorse in the field, and gained renewed favour with the recent publication by John Mullahy and Edward Norton, who espouse the use of such by recommending "…using a non-transformed dependent variable, such as a two-part model, untransformed linear regression, or Poisson" (Mullahy & Norton, 2023). Unlike competitor approaches, such 2PMs appear to be less susceptible to specification issues and functional form and distributional assumptions.

### 18.2.2 Modelling of Healthcare Costs (2000s-)

The RAND HIE made the modelling of individual healthcare costs a particular focus for health econometric research. Subsequently the literature broadened and models of individual healthcare costs have been used to parameterise decision models in cost-effectiveness analyses, to adjust for healthcare need in resource allocation in public healthcare systems, to inform risk adjustment in insurance-based systems, and to assess the costs attributable to factors such as smoking and obesity. As discussed above, the characteristics of healthcare costs pose substantial challenges for econometric modelling. They are non-negative, highly skewed and heavy tailed and often exhibit a large mass point at zero. In addiction the relationship between covariates and costs may be nonlinear. The policy relevance and challenges of modelling healthcare costs led to the development of a wide range of econometric approaches that went beyond the use of log-transformed costs (Jones, 2012).

Much of the focus in the use of regression methods for the analysis of healthcare cost data has centred on predictions of the conditional mean of the distribution. In the 2000s attention shifted away from linear regression models to semiparametric and flexible parametric estimators. A popular semiparametric approach was generalized linear models or GLMs (e.g., Manning & Mullahy, 2001; Buntin & Zaslavsky, 2004; Manning, Basu & Mullahy, 2005; Manning, 2006). This framework offered a relatively simple way to incorporate non-linearities in the relationship between the conditional mean and observed covariates. Furthermore, GLMs allow for heteroskedasticity through a choice of a distribution, which specifies the conditional variance as a function of the conditional mean. GLMs use pseudo-maximum likelihood estimation where the researcher is required only to specify the form of the conditional mean and variance.

In a conventional GLM the choice of link and distribution has to be specified a priori. In practice the most frequently used GLM specification for medical costs has been the log-link with a gamma variance (Blough, Madden & Hornbrook, 1999; Manning & Mullahy, 2001; Manning et al., 2005). Basu and Rathouz (2005) developed a flexible semiparametric approach to the problem of selecting the appropriate link and variance functions. Their extended estimating equations estimator (EEE) approach used a Box-Cox transformation for the link function and either a power variance or quadratic variance function for the distribution. As such, the functional forms of the link and distribution are estimated from the data.

While the GLM framework has attractive properties for researchers concerned with predictions of the conditional mean, there are important limitations with this method. GLMs have been found to perform badly with heavy-tailed data (Manning & Mullahy, 2001), and they implicitly impose restrictions on the entire distribution. For example, whatever distribution is adopted, the skewness is directly proportional to the coefficient of variation, and the kurtosis is linearly related to the square of the coefficient of variation (Holly, 2009). While they may be well placed to estimate the conditional mean, they cannot produce estimates of the full conditional distribution including conditional tail probabilities.

While the mean is an important feature of a distribution, which is essential when the analysis is concerned with the expected total cost, it is not the only aspect that may be of interest to policymakers (Vanness & Mullahy, 2006). Analysis based solely on the mean misses out potentially important information in other parts of the distribution. A growing literature in econometrics has developed techniques to model the entire conditional distribution, thus 'go beyond the mean'. In health economics, there is a particular emphasis on identifying individuals or characteristics of individuals that lead to very high costs.

Other approaches were developed that offered greater flexibility in terms of their potential applications, imposing fewer restrictions on skewness and kurtosis and allowing for a greater range of estimated effects of a covariate. These include the use of flexible parametric distributions for modelling healthcare costs (Manning et al., 2005; Jones, Lomas & Rice, 2014), which have been applied to healthcare costs principally in order to overcome the challenge posed by heavy-tailed data. Unlike the GLM framework, these models assume a functional form for the entire distribution with estimation by maximum likelihood. A related development is the use of finite mixture models (FMM), which allow the distribution to be estimated as a weighted sum of distribution components (Deb & Trivedi, 2006). These are also estimated using maximum likelihood but are often referred to as semiparametric, because the number of components could, in principle, be increased to approximate any distribution.

Other developments in the estimation of healthcare costs have been less parametric, typically involving dividing the outcome variable into discrete intervals and estimating parameters for each of these intervals. Gilleskie and Mroz (2004) proposed using a conditional density approximation estimator, with the density function approximated by a set discrete hazard rates. To implement this, Jones, Lomas and Rice (2015) used an approach based on that of Han and Hausman (1990), by creating a categorical variable that denotes the cost interval into which each observation falls and running an ordered logit with this as the dependent variable. This tied into a related literature on semiparametric estimators for conditional distributions (Han & Hausman, 1990; Foresi & Peracchi, 1995; Chernozhukov, Fernández-Val & Melly, 2013).

## 18.3 Models for Qualitative and Categorical Dependent Variables (1980s-)

A major challenge in empirical health economics, recognised in the RAND HIE and expanded upon from the 1980s onwards, is the qualitative nature of many dependent variables of interest and a requirement for models to analyse binary, ordered, and multinomial outcomes.

Binary dependent variables are among the most common in health economics. They arise whenever the outcome of interest is dichotomous, such as whether an individual utilizes healthcare services, purchases health insurance, or engages in health-related behaviours like smoking cessation. In such cases, the econometric problem is to model the probability of participation as a function of observed characteristics. While the linear probability model provided a simple approximation, it suffered from well-known drawbacks, including heteroscedasticity and the possibility of predicted probabilities outside the unit interval. As a result, nonlinear specifications such as the probit and logit models, were typically preferred. These models were often motivated by a latent variable framework in which the observed binary outcome reflects an underlying propensity.

Applications of binary response models in health economics have been extensive. For example, Buchmueller and Feldstein (1997) analysed the decision of employees to switch health insurance plans following changes in employer contributions. Using a probit model, they showed that price changes in premiums have significant effects on switching behaviour. Such studies highlight the importance of price sensitivity in insurance markets and provide insights into policy reforms affecting coverage choices. More generally, binary models have been used to analyse physician visits, hospital admissions, preventive care uptake, and risky health behaviours (Urban, Anderson & Peacock, 1994; Mainous & Gill, 1998; Evans, Farrelly & Montgomery, 1999, among many others). Their flexibility and interpretability along with the ubiquitous use of variables measured as binary outcomes have made them indispensable tools in applied health econometrics.

Many health outcomes that have been the focus of empirical study are inherently ordinal rather than purely binary. A prominent example is self-assessed health status, which is often reported in categories such as excellent, good, fair, poor, and very poor. Ordered probit models became the favoured approach to exploit the ordinal structure of such data, where the observed categorical outcome was assumed to arise from a latent continuous variable that is partitioned by threshold parameters. This framework allowed researchers to estimate the extent to which explanatory variables shift the distribution of the latent health index, while respecting the ordinal nature of the observed categories.

Applications of ordered models in health economics were widespread. Kenkel (1995) used ordered probit models to analyse self-reported health and activity limitations, while Chaloupka and Wechsler (1997) applied similar methods to study cigarette consumption categories. Levinson and Ullman (1998) employed an ordered probit to examine the adequacy of prenatal care, and Theodossiou (1998) used related

ordered logit models to investigate mental distress, finding significant effects of unemployment on mental health outcomes.

The prominence of self-assessed measures of health on an ordered scale due largely to the low cost of collecting such data in surveys, while initiating research interest led to significant concerns over measurement error and differences in reporting behaviour, which are assumed to be systematic and linked to preferences and tastes varying across both individuals and population groups. This inspired innovation in applications of the ordered probit model to cater for such reporting bias. We discuss these in Section 18.4 below as their development was heavily linked to the increasing availability and use of cross-sectional and panel data surveys.

Where outcome categories are delineated by known cut-points, for example, by income level ranges, the related *Interval Regression* is more suitable and not subject to criticisms aimed at ordered models. Sutton and Godfrey (1995) applied this approach to analyse alcohol consumption categories, while Donaldson, Jones, Mapp and Olson (1998) suggested its use in modelling willingness-to-pay data obtained from categorical scales. Although, based on solid foundations, the technique's popularity and hence application in health econometrics has been dependent upon the availability of appropriate interval data.

### 18.3.1 Multinomial Choice Models

In many health economics applications, individuals face a choice among multiple discrete alternatives. Examples include the selection of health insurance plans, the choice of healthcare provider, and treatment decisions. Multinomial choice models have been popular to handle such unordered categorical outcomes. The multinomial logit model has been the most widely used specification, often derived from a random utility framework. In this setting, individuals choose the alternative that maximizes their utility, which depends on both observed and unobserved characteristics. The resulting choice probabilities take a convenient closed form, under the assumption of independently distributed extreme value errors.

Applications of multinomial logit models in health economics have been widespread. For example, Dowd, Feldman, Cassou and Finch (1991) used such models to analyse insurance plan choice, while Haas-Wilson and Savoca (1990) examined the selection of healthcare providers. These models allow researchers to assess how individual characteristics and plan attributes influence choice behaviour. A key limitation of the multinomial logit model, however, is the independence of irrelevant alternatives ($IIA$) assumption, which implies that the relative odds of choosing between two alternatives are unaffected by the presence of other options. This assumption is often unrealistic in healthcare settings, where alternatives may be close substitutes.

To address this limitation, researchers adopted more flexible models. The nested multinomial logit model allows for correlation among alternatives within groups. Gertler, Locay and Sanderson (1987) applied this framework to study the demand for medical care in Peru, modelling both the decision to seek care and the choice

of provider. Feldman, Finch, Dowd and Cassou (1989) used a nested structure to analyse health insurance choices, showing that plan selection is sensitive to out-of-pocket costs and that ignoring substitution patterns may lead to misleading conclusions. A further alternative is the multinomial probit model, allowing for a general covariance structure among error terms. Although computationally more demanding, advances in simulation methods in the 1990s made estimation feasible. Börsch-Supan, Hajivassiliou and Kotlikoff (1992) and Hoerger, Picone and Sloan (1996) used multinomial probit models to study living arrangement choices among the elderly, while Bolduc, Lacroix and Muller (1996) applied the model to provider choice in developing countries. These studies demonstrated that relaxing restrictive assumptions can lead to substantially different estimates of price responsiveness and substitution patterns.

### 18.3.2 Models for Count Data

A further area of focus in empirical health economics that emanated from the RAND HIE and the development of the 2PM was the consideration of count data models. This is an area where methodological developments pioneered by applications to health data have diffused into the broader econometrics literature.

Recognising the discrete integer valued and non-negative nature of many health-related outcomes, these models are particularly suited to analysing measures of healthcare utilization, such as the number of physician visits, hospital admissions, or prescriptions filled over a given period. In health economics, such variables are often characterized by a highly skewed distribution, with a large proportion of zeros and a long right tail. This reflects the empirical reality that many individuals may not use healthcare services within a given period, while a smaller group accounts for a large number of visits or treatments. The distributional features of these outcomes pose challenges for standard linear regression models, which are ill-suited to handle discreteness, non-negativity, and skewness.

Empirical applications are abundant. Two stalwarts of count data analysis, Colin Cameron and Pravin Trivedi, analysed multiple measures of healthcare utilization using data from the 1977–78 Australian Health Survey, which became a benchmark dataset in this literature (Cameron & Trivedi, 1986). Subsequent studies using the same data include Cameron, Trivedi, Milne and Piggott (1988), who examined hospital admissions and medication use, and Cameron and Windmeijer (1996), who compared goodness-of-fit measures across count data models. Other applications include Cauley (1987), who estimated models for outpatient visits in the Kaiser Permanente system, and Pohlmeier and Ulrich (1995), who analysed physician visits in Germany. Deb and Trivedi (1997) extended this framework to multiple utilization measures among the elderly using U.S. survey data.

While the standard model for count data is Poisson regression, early reliance on this approach was regarded as insufficient. The assumption of equidispersion (equality of mean and variance) was rarely satisfied in heath care data. Healthcare

utilization data are typically overdispersed, with the variance exceeding the mean, and it was recognised that such overdispersion likely arose from unobserved heterogeneity across individuals, such as differences in health status, preferences, or access to care. Accordingly, the negative binomial model, assuming that the Poisson mean is a random variable to capture unobserved heterogeneity, became a popular alternative. For example, Vistnes and Hamilton (1995) used a negative binomial specification to analyse mothers' demand for paediatric care, while Grootendorst (1997) applied similar methods to study pharmaceutical utilization among older individuals.

However, healthcare utilization data often exhibit excess zeros or more complex forms of heterogeneity. In an attempt to capture these features two broad approaches were adopted. To address the excess zeros problem, hurdle models and zero-inflated models were applied. Pohlmeier and Ulrich (1995) applied hurdle models to physician visits, demonstrating that separating participation and intensity decisions improves model fit. Similarly, Arinen, Sintonen and Rosenqvist (1996) compared single-equation and two-part models for dental visits, highlighting the importance of accounting for the large mass of zeros. Numerous other applications of the two-part or hurdles model exist in the literature, for example, Álvarez and Delgado (2002), Chang and Trivedi (2003), and Sarma and Simpson (2006), showing how different factors influence the decision to access care and intensity of care conditional on access. Zero-inflated models provided an alternative by assuming that the data are generated from a mixture of two processes: one generating structural zeros and another generating counts (including zeros) (Mullahy, 1986). These models were particularly useful when some individuals are not at risk of using healthcare services, while others who are at risk do not access care. The two most prominent specifications were zero-inflated Poisson and zero-inflated negative binomial models. Examples include analysis of physician visits (Yen, Tang & Su, 2001; Sarma & Simpson, 2006), the number of pharmacy visits (Chang & Trivedi, 2003), and prescriptions (Street, Jones & Furuta, 1999) and the number of cigarettes consumed (Sheu, Hu, Keeler, Ong & Sung, 2004; Bauer, Göhlmann & Sinning, 2007).

Concerns about the influence of unobserved heterogeneity in count data models led to more flexible modelling approaches being developed. For example Mullahy (1997) emphasised that unobserved heterogeneity may be correlated with key variables such as insurance coverage, Pohlmeier and Ulrich (1995) argued that supply-side factors, such as provider availability, may influence utilization but are often unobserved in survey data. In an attempt to reflect better such heterogeneity Gurmu (1997) proposed a semiparametric approach using Laguerre series expansions to approximate the distribution of unobserved heterogeneity and which illustrated improved consistency and fit. Cameron and Johansson (1997) introduced an approach based on polynomial expansions around a Poisson baseline which allowed for deviations from the Poisson distribution in both the mean and variance, offering an alternative to the negative binomial model, particularly in cases of under-dispersion. These developments highlighted increasing sophistication of count data models and their ability to capture complex features of healthcare utilization.

An important strand in the evolution of count data models was the introduction of mixture models. Deb and Trivedi (1997) extended the standard framework by

developing a finite mixture negative binomial model. Two points of support were found to adequately describe the distribution of unobserved heterogeneity, suggesting two latent populations, the 'healthy' versus the 'ill'. Model selection criteria provided support in favour of the mixture model over both hurdle and negative-binomial specifications when applied to the demand for medical care by the elderly in the US. Further evidence in support of mixture models over hurdle models in the context of utilisation data was provided by Deb and Trivedi (2002) demonstrating how latent class models can be used to identify groups, distinguishing between infrequent and frequent users of care, using data on counts from the RAND HIE. They showed that policy relevant estimates of price elasticities of healthcare, and probabilities of high-use deviate substantively across the latent class and hurdle models. Finite mixture models have also been used in evaluating health interventions. Conway and Deb (2005) showed that accounting for latent heterogeneity between different types of pregnancies reveals significant effects of prenatal care on birth weight that are obscured in more conventional approaches. This highlights the importance of properly modelling heterogeneity when assessing policy effectiveness.

Building on the literature on the relative merits of the two-part, mixture and multiple spell models (an extension of count data models to multiple spells of contact with health services, Santos Silva & Windmeijer, 2001), Winkelmann (2004) proposed a two-part, or hurdle, model where rather than specifying a negative binomial distribution for the positive part of the distribution, they used a Poisson-log-normal distribution with individual-specific random effects to capture to capture heterogeneity in utilisation patterns. Evaluating healthcare reforms using data from the German Socio-Economic Panel on general practice and specialist visits, they demonstrated a superior fit of the model over finite mixtures, and hurdle models using negative binomial distributions. A key message from this literature is that model specification is data and context dependent and that flexibility in modelling approach to count data is recommended.

Recently Wooldridge (ref needed) has stressed the quasi (or pseudo) maximum likelihood estimator (QMLE) property of the Poisson model, which holds when the model is estimated by maximum likelihood with robust standard errors. This property ensures consistent estimates, so long as the Poisson conditional mean is correctly specified. It holds even if the conditional variance is misspecified - as it often will be with health data. This QMLE property is not shared by alternatives such as the negative binomial model, which limits their robustness and appeal. However, first order QMLE relies on the correct specification of the conditional mean. With health data, alternative models that have mean functions that differ from the exponential condition mean function of the Poisson model - such as hurdle, zero-inflated or finite mixture models - may be correctly specified when the Poisson models is not. Even if attention is restricted to models that have an exponential conditional mean (a "log link" in the GLM framework), the finite sample performance of the estimator may be improved by using a different conditional variance - such as using the gamma rather than the Poisson variance. This is because different variance functions give different weighting to extreme observations in the likelihood function. More generally, this

ties in with Holly's (2009) work on higher order pseudo/quasi maximum likelihood methods.

### 18.3.3 Duration Models

While count data measures the number of events that occur during a fixed interval, survival analysis focuses on the time elapsed between events. The analysis of survival and duration data has played a role in the development of econometric methods in health economics, reflecting the importance of timing in many health-related outcomes. Early applications in health economics highlighted the usefulness of survival analysis for understanding behavioural and institutional dynamics. Researchers have applied these models to a wide range of topics, including mortality risk (Behrman, Sickles & Taubman, 1990; Behrman, Sickles, Taubman & Yazbeck, 1991), initiation of smoking (Douglas & Hariharan, 1994; Douglas, 1998), duration of hospital stays (Hamilton & Hamilton, 1997), or time until first use of health services (Philipson, 1996).

Both parametric and semiparametric approaches to duration modelling have been adopted. Weibull models have been a popular choice because they can capture increasing or decreasing hazard rates over time—a feature that is particularly relevant in health applications, where the risk of an event may rise with age or decline as individuals recover from illness. However, a variety of alternative approaches, including the exponential, log-normal, log-logistic, and generalized gamma models have been developed to capture the shape of the hazard function, including non-monotonic patterns that are often observed in empirical health data. Applications of parametric duration models in health economics include Morris, Norton and Zhou (1994); Norton (1995).

In historical context, a major methodological advance in duration modelling applied to health data was the adoption of semiparametric approaches, most notably the Cox proportional hazards model. This represented an important shift away from fully parametric models, by placing less assumptions on the underlying baseline risk over time while retaining a clear interpretation of covariate effects. Despite its advantages, the Cox model relies on the proportional hazards assumption, which may not always hold in practice. This limitation led to further methodological developments, including stratified versions of the model that allow different groups to have distinct baseline hazard functions. Such extensions have been particularly useful in health economics, where heterogeneity across populations is often substantial. Applications of semi-parametric duration models in this field include Behrman et al. (1990, 1991); Philipson (1996).

As the field progressed, researchers increasingly recognized the importance of unobservable heterogeneity, often referred to as frailty, in shaping duration outcomes. For example, populations may appear to exhibit declining hazard rates over time simply because individuals with higher risks experience events earlier, leaving a healthier group behind. This insight led to the incorporation of latent heterogeneity into duration models. Various approaches were adopted to address this issue. Some models assumed

a specific distribution for the unobserved component, while others adopted more flexible, non-parametric representations, such as discrete mixtures. These methods allowed researchers to better capture the complexity of real-world health processes and to obtain more reliable estimates of duration dependence. Empirical applications examining unobservable heterogeneity in mortality risk include Behrman et al. (1990, 1991). Discrete time versions of the mixed proportional hazard model have been the core method in studies on the impact of health shocks on labour market outcomes such as early retirement (e.g., Disney, Emmerson & Wakefield, 2006).

In many contexts, events are not singular but involve competing risks, such as different causes of death or alternative pathways out of employment due to health conditions. Competing risk models allowed for the simultaneous analysis of multiple hazard processes, providing a richer understanding of the underlying dynamics. In addition, researchers adopted models that accommodate multiple spells, recognizing that individuals may experience repeated episodes of illness, recovery, or healthcare utilization over time. These extensions were often implemented within Markov or semi-Markov frameworks and built on earlier developments such as the Cox model. By incorporating multiple transitions and repeated events, they offered a more realistic representation of health and labour market dynamics. Applications of such competing risk models in health economics included Butler, Anderson and Burkhauser (1989); Hamilton and Hamilton (1997).

## 18.4 Social Surveys and Microeconometric Models (1990s–2000s)

Investments in micro-level survey datasets throughout the 1990s and beyond initiated an expansion in the type of research questions that could be addressed in health economics, particularly around socio-economic drivers of health, health inequalities, the links between health and other human capital investments such as education, and health and labour market experiences. The availability of rich household-level panel data expanded the empirical toolkit to address unobserved heterogeneity and explore dynamic behaviour. While many of the methods described in the preceding sections remained relevant for empirical health economics, the availability of panel data necessitated new models accounting for individual-specific heterogeneity. We describe below key developments that resulted from the increased availability of survey data.

### 18.4.1 Self-Reported Health and Reporting Behaviour

Subjective health measures, particularly self-assessed health (SAH), became a central component of empirical work in health economics, reflecting both increased data availability and methodological developments. Their widespread use was closely linked to the expansion of large-scale household surveys which collected health

information in a consistent and longitudinal format. Typically recorded as an ordered categorical variable—ranging from excellent to poor health—SAH provided a convenient and flexible proxy for overall health status in econometric analysis (Jones, 2009), and became a standard outcome variable in studies of health dynamics, labour market behaviour, and socioeconomic inequalities.

An attraction of subjective health measures is their ability to capture multiple dimensions of health within a single indicator. Unlike many objective measures, which focus on specific conditions or clinical markers, SAH reflects a broader assessment that includes physical, mental, and functional aspects of well-being. This made it particularly useful in applied work, where comprehensive objective measures were often unavailable. Empirical evidence demonstrated SAH to be a strong predictor of important outcomes such as mortality and morbidity (Idler & Benyamini, 1997), reinforcing its value as a summary measure. However, it was also clear that individuals may misreport their health status due to differences in access to medical information, healthcare utilisation, or awareness of underlying conditions. Evidence from studies linking survey responses to administrative records suggested the presence of both under-and over-reporting of health conditions, with reporting errors often being correlated with socioeconomic characteristics (Baker, Stabile & Deri, 2004, also see Stoye & Zaranko, 2020). In addition, the literature highlighted the possibility of 'justification bias', whereby individuals adjust their reported health to rationalise labour market outcomes or benefit receipt (Kerkhofs & Lindeboom, 1995; Kreider, 1999).

A further challenge arises from so-called reporting heterogeneity. Individuals with similar underlying health may report different categories due to differences in expectations, cultural norms, or reference groups; often referred to as differential item functioning (DIF). As a result, observed differences in SAH may reflect not only true variation in health but also differences in reporting behaviour, complicating both interpretation and cross-group comparisons.

In response to these challenges, the health economics literature has developed a range of methods to improve the use of subjective health measures. Important extensions relate to the fact that these ordered models have fixed threshold/boundary, parameters, and authors have relaxed these assumptions, often depending on observed characteristics; see, for example, Kerkhofs and Lindeboom (1995) and Pudney and Shields (2000). Although yielding more flexible approaches they failed to find much favour in the empirical literature as one typically needs to allocate observed characteristics to either reporting behaviour, or to the underlying construct of interest, say health: which often proved hard to justify.

#### 18.4.1.1 Survey Vignettes

In the 2000s researchers found that they could use responses to a set of hypothetical *survey vignettes* to so-call 'anchor' individuals' subjective responses to a common scale (King, Murray, Salomon & Tandon, 2004). The vignettes would typically describe the health status of a hypothetical individual and respondents would rate

this individual's health along with their own. Unlike previous approaches, this does not require partitioning characteristics into health-outcome and reporting-behaviour buckets. For example, Soloman, Tandon and Murray (2004); Bago d'Uva, Van Doorslaer, Lindeboom and O'Donnell (2008); Grol-Prokopczyk, Freese and Hauser (2011); Vonkova and Hullegie (2011); Peracchi and Rossetti (2012) used vignettes to model self-reported data on health status; Van Soest, Delaney, Harmon, Kapteyn and Smith (2011) for healthy behaviours; Rice, Robone and Smith (2012); Sirven, Santos-Eggimann and Spagnoli (2012) for satisfaction with health system performance; and Angelini, Cavapozzi and Paccagnella (2011); Kapteyn, Smith and Van Soest (2007); Paccagnella (2011) when studying work disability. The approach is only valid under the two identifying assumptions of response consistency (RC) and vignette equivalence (VE). RC assumes that individuals use the same mapping from the underlying latent scale to the available response categories when assessing the self-assessment as they use when assessing the corresponding vignettes. VE assumes that 'the level of the variable represented by any one vignette is perceived by all respondents in the same way and on the same unidimensional scale' (King et al., 2004). This implies that respondents agree on the underlying latent level of the concept under scrutiny, as depicted by the hypothetical situation described by the vignette, except for random error.

Mostly the empirical literature has attempted to investigate the validity of these assumptions, based on more rule-of-thumb approaches; and quite often both were found to fail. Both Peracchi and Rossetti (2013) and Greene, Harris, Knott and Rice (2021) have developed direct tests for these. In practical applications, however, these mixed findings across studies suggests that whether they are tenable assumptions varies significantly across surveys, subgroups, instruments of interest and the particular vignettes used. Despite a surge in interest in the early 2010s, coupled with the increased costs associated with included vignettes within sample surveys has led to the approach failing to maintain appeal. Despite their limitations, subjective health measures though, still remain very relevant outcomes, although their importance has declined as more objective alternatives, such as biomarkers and administrative data, have become more readily available (Benzeval, Kumari & Jones, 2016).

### 18.4.2 Panel Data Approaches

The increased availability of panel data, particularly household panels, from the 1990s onwards marked a significant shift in approaches to controlling for unobserved heterogeneity in a systematic way. This represented an advance over earlier cross-sectional studies, enabling more credible identification of causal relationships. Applied health economics research at the time focused on two key challenges: unobserved individual heterogeneity correlated with explanatory variables, and the use of nonlinear models to handle qualitative or limited dependent variables. Together, these issues shaped the development of econometric methods in the field, particularly in

the analysis of longitudinal data where dynamic elements such as persistence and behavioural adjustment are central.

Early applications in health economics demonstrated how fixed effects approaches could eliminate time-invariant unobserved factors that might otherwise bias results. For example, Lindeboom, Portrait and Van den Berg (2002) analysed cognitive status and emotional well-being using fixed effects models, showing how major life events such as bereavement influence mental health outcomes among older individuals. This type of work illustrated how panel methods made it possible to isolate the impact of changes within individuals over time. The same methodological logic has been applied to the study of healthcare providers. For example, Carey (2000) examined hospital behaviour by analysing the relationship between length of stay and costs, accounting for unobserved institutional characteristics such as quality of care. By allowing these factors to be correlated with observed variables, the analysis provided more reliable estimates of cost relationships. The development of two-way fixed effects models further strengthened the empirical toolkit. By incorporating both individual and time effects, these models became particularly useful for evaluating policy interventions. M.-C. Lee and Jones (2004), for instance, used this approach to study the effects of global budgeting in Taiwan, exploiting within-provider variation to identify behavioural responses. Their findings highlight how providers adjusted service composition in response to policy changes.

The study of health inequality has been a central pillar of health economics, driven by evidence that health disparities persist and often widen despite policies promoting equal access and social inclusion (Jones, Rice, Smith, Ginnelly & Sculpher, 2005). A primary tool for measuring socioeconomic inequality in health has been the *health concentration index*, derived from the health concentration curve (Wagstaff, Van Doorslaer and Paci (1989)). While measuring and comparing, across time or jurisdictions (often countries), was the initial use of inequality indices, interest in understanding the drivers of inequality soon followed. Wagstaff, Van Doorslaer and Watanabe (2003) proposed that the concentration index could be decomposed by factors into component parts consisting of an explained component, comprising contributions from variables like income, education, and activity status, and an unexplained component. This approach has been superseded by subsequent developments in the literature on decomposition analysis but, at the time, it provided a catalyst for econometric analysis to implement the decompositions (Van Doorslaer & Koolman, 2004). Newly available longitudinal data allowed for the analysis of health *mobility* and long-run inequality. Shorrocks (1978) provided early indices for income mobility that have been adapted to health. Research by Hauck and Rice (2004) found evidence of substantial mobility in mental health but noted that persistence (low mobility) is more common among lower-income and less-educated groups.

#### 18.4.2.1 Spatial Health Models

As data availability improved, linear panel models were extended to incorporate *spatial dimensions*. This reflected growing recognition that health outcomes and

healthcare provision or expenditures are often influenced by peer or geographic spillovers (see Baicker, 2005; Moscone, Tosetti & Vittadini, 2011). These interactions between individuals or providers of care result in behaviours and decisions that empirically translate into important structural correlations. Empirical interpretations typically result in the specification of a spatially lagged dependent variable model where the relationship between the cross-section observations (e.g., providers of care) is expressed by the use of a non-negative spatial weights matrix, $W$, with elements $w_{ij}$ representing the weight of the relationship between units $i$ and $j$. If this weight is non-zero (i.e., $w_{ij} \neq 0$) units $i$ and $j$ are spatially related (considered a neighbour). The specification of $W$ has usually been based on some measures of distance, using for example the inverse of the distance between healthcare providers, with diagonal elements set to zero since an observation is not a neighbour to itself. Typically the weights for each row are standardised to sum to unity, which facilitates interpretation of results (see Anselin, 1988). Estimation of spatial models have focused on the use of Instrumental Variables (IV), Maximum Likelihood Estimation (MLE) and Generalized Method of Moments (GMM) approaches.

In an early example of the application of spatial models in health economics, Mobley (2003) examined hospital market competition in California. Recognising that agents might be making strategic choices based on the decisions of neighbouring agents, hospital prices were modelled using a spatial econometric specification which allowed the for prices of rival hospitals, based on spatial proximity, to have spillover effects. This illustrated how interdependence can be interpreted as an agent's choices being directly impacted by the choices made by neighbours, as distinct from situations where similar behaviours arise due to common neighbourhood characteristics. Given that a hospital is likely to compete with many other hospitals in the local market weights were based on a combination of distance and the $k$ closest neighbours. In a similar work, Gravelle, Santos and Siciliani (2014) explored hospital competition where prices are regulated, extending these ideas to consider whether the quality of hospital care is influenced by the quality of hospitals in proximity. The approach departed from the usual way to test whether competition affects hospital quality which typically relied on exploring the relationship between measures of quality (for example, hospital mortality) and measures of competition such as the Herfindahl index or the number of rival hospitals. Instead a spatial lag model was specified where the quality of a hospital was determined by measures of market structure together with the quality of rival hospitals weighted by distance within a 30-minute travel time. The set of measures of market structure were similar to those used in non-spatial specifications and included conventional measures of competition which were often the focus for testing for evidence of competition on quality. By contrast, in the spatial specification, interest focused on the sign of the spatial lag to test whether the quality of care of rivals were strategic complements or substitutes.

The use of spatial health models is health econometrics has been motivated by a desire to test economic theories predicting interactions within networks at a micro-level, for example hospital organisation and quality responding to local competition and practice (Moscone et al., 2011; Baltagi & Yen, 2014; Longo, Siciliani, Gravelle & Santos, 2017), or a macro-level, for example decision making

and spillovers across local health authorities and regions (Costa-Font & Pons-Novell, 2007; Moscone, Knapp & Tosetti, 2007; Ortiz & Masiero, 2013; Atella, Belotti, Depalo & Piano Mortari, 2014). Extensions to count data and hurdle models have also been considered (Neelon, Ghosh & Loebs, 2011). These approaches have illuminated the importance of considering the inter-connectedness across economic agents. Baltagi, Moscone and Santos (2018) provide an accessible introduction to the use of spatial models in health economics, while Moscone and Tosetti (2014), provide a more technical exposition.

Taken together, the above developments illustrated how linear panel models evolved into a flexible and widely used framework in health econometrics, capable of addressing increasingly complex data structures while maintaining straightforward interpretation. It is worth noting that random effects models, unlike in other social science and bioscience disciplines, rarely feature in linear models in health econometrics. Despite specification tests of fixed versus random effects such as that provided by Hausman (1978), the wealth of available data in household panels rarely leads to gains in efficiency allowed by a random effects specification when consistently estimated, to outweigh concerns over endogeneity bias brought about by unobserved heterogeneity.

#### 18.4.2.2 Dynamic Panel Models

The recognition that many health-related outcomes exhibit persistence over time led to the adoption of *dynamic panel data models*. In contrast to static specifications, such models explicitly accommodate the influence of past outcomes on current behaviour, capturing processes such as habit formation, adjustment, and state dependence. Introducing lagged dependent variables raises important econometric challenges, however, particularly due to their correlation with unobserved individual effects. A key step in the econometrics literature was the development of Generalised Method of Moments (GMM) estimators by Arellano and Bond (1991), which provided a practical solution by using lagged variables as instruments. Subsequent refinements, including System GMM incorporating both levels and differences in lags as instruments (Arellano & Bover, 1995; Blundell & Bond, 1998), improved efficiency and expanded the applicability of these methods. These techniques were widely adopted in health economics. For example, Brown, Coffman, Quinn, Scheffler and Schwalm (2006) used dynamic panel methods to analyse the impact of Health Maintenance Organizations on physician supply, capturing persistence in provider behaviour and addressing endogeneity concerns. Similarly, Jones and Labeaga (2003) applied these methods to study tobacco consumption within a rational addiction framework, demonstrating how failure to account for dynamics and heterogeneity leads to misleading conclusions. Additional applications, such as Tamm, Tauchmann, Wasem and Greß (2007), highlighted the usefulness of dynamic models in analysing market behaviour in health insurance. The adoption of dynamic panel methods represented a major step in the evolution of health econometrics, allowing researchers to better capture behavioural processes that unfold over time.

#### 18.4.2.3 Nonlinear Panel Models

Section 18.2 highlighted the role of outcomes that are inherently qualitative or categorical. Variables such as self-assessed health status, labour force participation, doctor visits, and hospital utilisation did not fit naturally within the classical linear framework, and their analysis drove the adoption of categorical and count data models. Over time, these models were embedded within panel data settings, allowing researchers to account for persistence, unobserved heterogeneity, and dynamic behavioural responses (Munkin & Trivedi, 1999; Riphahn, Wambach & Million, 2003; Van Ourti, 2004). Of particular importance in this strand of literature, given the focus on nonlinear models and the incidental parameter problem encountered with the specification of fixed effects in such contexts, is the parameterisation of unobserved individual-specific heterogeneity through approaches set out in Mundlak (1978) and Chamberlain (1980), and formalised as the Correlated Random Effects (CRE) models by Wooldridge (2010), and their extensions to incorporate time effects in unbalanced panels (see Wooldridge, 2019).

A key development in nonlinear panel data models was the incorporation of dynamics through lagged dependent variables, typically within random effects probit or ordered probit frameworks. Importantly this allowed the differentiation between true state dependence and persistent unobserved differences across individuals. Empirical studies have consistently found strong intertemporal correlation in measures such as self-assessed health. Adopting a CRE approach, Contoyannis, Jones and Rice (2004) showed that self-assessed health exhibits substantial positive state dependence, and also demonstrated that the apparent role of unobserved heterogeneity diminishes once initial conditions and within-individual averages of explanatory variables are properly accounted for. This line of work reflected a broader shift in the literature towards more careful treatment of dynamic processes in discrete choice models.

A central methodological issue in this context is the treatment of *initial conditions*. Because the first observed outcome may itself be influenced by prior unobserved factors, failure to address this problem can lead to biased estimates of state dependence. The approach proposed by Wooldridge (2005), of conditioning on the initial observation and explanatory variables, became particularly influential. Its practical appeal led to widespread adoption in empirical studies. For example, Gannon (2005) applied a dynamic panel probit model to analyse labour force participation decisions, explicitly incorporating health limitations and past behaviour. Similarly, Nolan (2007) used a dynamic random effects Poisson model to study general practitioner visits, again relying on the Wooldridge approach to handle initial conditions. These applications illustrate how state dependence in behaviour can be captured while controlling for unobserved heterogeneity using easy-to-implement panel data models.

Alternative solutions to the initial conditions problem have also been explored. Heckman's earlier approach, for example, remains relevant in certain contexts and was employed by Arulampalam and Bhalotra (2006) in a Markov model of infant mortality. The existence of multiple methodological strategies highlights the importance of modelling choices in empirical work, particularly when both state dependence and heterogeneity play significant roles. Overall, panel data models evolved into

a flexible and powerful framework, though their implementation often requires strong assumptions. They offer rich insights into persistence, heterogeneity, thereby contributing to a deeper and more nuanced understanding of health and healthcare dynamics.

### 18.4.3 Multivariate Models and Causal Effects

In models of the demand for health and models used to construct health status indexes, problems of endogeneity and selection bias were complicated by the fact that the key outcome, health, is inherently unobservable. An early approach to dealing with this was to allow health to be proxied by indicator variables. The Multiple Causes-Multiple Indicators, or MIMIC, model was adopted by health economists to deal with the problem of latent variables. In the early literature MIMIC models were estimated as LISREL (Linear Structural Relationships) models. Examples of the the use of LISREL models of the demand for health include Van de Ven and Van der Gaag (1982) and Wagstaff (1986). Wolfe and van der Gaag (1981) and Van Vliet and Van Praag (1987) used MIMIC models in the derivation of a health status indexes. Behrman and Wolfe (1987) estimated a structural model of health production functions for maternal and child health in Nicaragua.

Moving on from MIMIC models, nonlinear multivariate models bring together equations that are related through common unobservable factors. The aim is to control for the selection bias caused by these unobservables and thereby identify causal effects of relevant treatment variables. All of this is done in the context of nonlinear microeconometric models with qualitative or limited dependent variables. The computational challenge of specifying the joint distribution of multiple outcomes and multiple treatments has been handled by methods such as maximum simulated likelihood (MSL), Bayesian MCMC, and the use of copulas. In a series of papers Pravin Trivedi and co-authors modelled healthcare expenditures and utilisation along with multinomial choices of insurance coverage, while allowing for unobservables that influence both the choice of insurance plan and the use of healthcare. For example, Deb and Trivedi (2006) assumed a parametric distribution for the latent factors and estimated the joint distribution by MSL. Zimmer and Trivedi (2006) use copulas to model the joint distribution by linking together the marginal distributions for the outcomes through a copula function. The Bayesian MCMC approach provided a convenient way of handling the computational challenge by systems of nonlinear equations, especially when the latent variables inherent in the models were handled as missing data using a data augmentation approach (e.g., Hamilton, 1999; Deb, Munkin & Trivedi, 2006).

## 18.5 Quasi-Experiments and Causal Inference (2000s–)

In recent decades causal inference has become the dominant concern of empirical health economics, reflecting the field's focus on policy evaluation and evidence-based decision making (Jones & Rice, 2012). While early empirical work was often descriptive, modern health econometrics places primary emphasis on identifying the effects of medical treatments, health technologies, and policy interventions on outcomes such as health status, healthcare utilisation, and costs. This shift has been driven by the need to inform resource allocation in health systems, where credible estimates of causal effects are essential for determining the effectiveness and efficiency of competing interventions.

Over recent decades there has been a shift towards a design-based approach. This favours being explicit about two things. First, pre-estimation specification of the estimand of interest – usually specified in terms of well-defined treatment effects, often focused on average outcomes. Second, the source of exogenous variation that is assumed to provide identification of these effects. This approach has seen a resurgence of linear regression models, often with fixed effects, estimated by least squares. This method offers transparency and familiarity and its advocates stress that it provides the best linear approximation to the conditional expectation function. It yields estimates of average treatment effects that are a weighted average of heterogeneous treatment effects. Regression provides a means of balancing observed covariates in a quasi-experimental setting.

The fundamental issue underlying this approach is the so-called evaluation problem, which arises from the fact that it is impossible to observe the same individual both with and without treatment at the same time. This insight, formalised in the potential outcomes framework, has played a key role in shaping modern econometric thinking. Each individual is understood to have multiple potential outcomes corresponding to different treatment states, but only one of these is observed in practice. The challenge for the researcher is therefore to construct a credible counterfactual — an estimate of what would have happened in the absence of the intervention. This conceptual framework has become a cornerstone of applied econometrics in health and beyond.

Randomized controlled trials (RCTs) represent an important benchmark in this context, as their use of random assignment ensures that treatment status is independent of both observed and unobserved characteristics. This property eliminates selection bias and allows for relatively straightforward estimation of causal effects. In health economics, RCTs have long been used in clinical research, particularly in the evaluation of new treatments and technologies. However, their application is often limited in broader policy contexts, where ethical considerations, high costs, and practical constraints make randomization difficult or impossible. As a result, much of the empirical work in health economics relies on observational data.

The reliance on observational data introduces the problem of selection bias, which has been a central focus in the development of econometric methods in the field. Selection bias arises when treatment assignment is correlated with factors that also influence outcomes, including both observable characteristics, such as age or income, and unobservable factors, such as preferences or underlying health risks. If these

factors are not properly accounted for estimated treatment effects may be severely biased.

Over time, a wide range of econometric techniques have been developed to deal with selection and to approximate the conditions of a controlled experiment. If selection operates only through observable characteristics, methods such as regression adjustment, matching, and reweighting can be used to construct comparable treatment and control groups. Applied economists are often sceptical of this approach to identification and are reluctant to rule out selection on unobservables. In more complex settings, where unobserved confounders are present, researchers have turned to quasi-experimental methods, including instrumental variables (IV), regression discontinuity designs (RDD) and, most commonly, difference-in-differences (DID) approaches. These methods exploit sources of exogenous variation in treatment assignment, allowing for more credible identification of causal effects.

In practice, the application of causal inference methods in health economics has involved a combination of conceptual and empirical steps. Researchers must first define the causal effect of interest, then identify an appropriate strategy to estimate it, and finally implement statistical methods using available data. Importantly, the emphasis is not only on testing theoretical predictions but also on producing policy-relevant estimates, such as average treatment effects, for specific populations. As a result, the credibility of empirical findings depends critically on the plausibility of the identifying assumptions and the robustness of the chosen econometric approach.

### 18.5.1 Instrumental Variables

The use of IV has a long history in health economics. A pioneering paper was Auster et al.'s (1969) analysis of death rates across the United States in 1960. They specified a Cobb-Douglas model for mortality rates, as a function of medical care and environmental variables. This was estimated by two-stage least squares (2SLS) to allow for the possible endogeneity of medical care, recognising that aggregate mortality rates may influence the level of spending on medical care at the state level. Acton (1975) used the linear simultaneous equations model to study the impact of nonmonetary factors on healthcare use. In an approach that was rooted in a structural model of household health production, Rosenzweig and Schultz (1983) addressed the problem of unobservable heterogeneity bias in the demand for child health inputs. For example, a mother who has a history of complications during previous pregnancies may be more likely to seek early prenatal care. Their proposed solution was to find instruments, such as input prices, that predicted the use of medical care but did not have an independent effect on health outcomes.

An important and influential application of IV in health economics, and more broadly to health services research, was the contribution of McClellan, McNeil and Newhouse (1994). The paper considered intensity of treatment for patients with acute myocardial infarction (AMI), such as catheterization and revascularization, on mortality. Recognising the selection problem that more intensive treatment is

provided to more severely ill patients, but not the most severely ill - as they are less likely to survive treatment - they employed distance to hospital (differential difference between the nearest hospital that treats AMI intensively and the nearest hospital) as an instrument. The study served to illustrate the importance of adjusting for selection effects when identifying treatment effects from observational data.

Instrumental variables methods have been applied across a range of topics in health economics, including the effects of insurance coverage, the impact of medical treatments, and the role of provider behaviour. In practice, identifying credible instruments has proven to be one of the most challenging aspects of empirical work in the field. Commonly used instruments when considering effective care include, distance to provider (for example, McClellan et al., 1994), provider congestion (for example Hoe, 2022; Godøy, Haaland, Huitfeldt & Votruba, 2024), provider's tendency to administer treatment (for example, Brookhart, Wang, Solomon & Schneeweiss, 2006), timing of admission (for example, Ho, Hamilton & Roos, 2000), and using ambulance assignment to examine hospital quality (for example, Doyle Jr, 2005). This demonstrates the breadth of applications and the creativity involved in identifying suitable instruments. At the same time, their successful use depends critically on the strength and plausibility of the chosen instruments, as well as careful interpretation of the resulting estimates. Critically, the success or otherwise of the use of instrumental variables largely lies in the credibility of the exclusion restriction and its relevance. Instruments used in the early literature are unlikely to be perceived as valid by more modern standards (see Rashad & Kaestner, 2004; French & Popovici, 2011 for discussions), and greater attention has been placed on investigating such assumptions, for example, through the use of falsification tests (for example, Dranove & Wehner, 1994).

#### 18.5.1.1 Genetics as IVs

The use of genetic variation as an IV has emerged as a significant development in causal inference within health economics, most prominently through the framework of Mendelian randomization (von Hinke Kessler Scholder, Smith, Lawlor, Propper & Windmeijer, 2011). This approach builds on the principle that genetic variants are randomly allocated at conception according to Mendel's laws, creating a form of natural experiment. Because genotypes are determined prior to birth and are generally independent of socioeconomic and behavioural confounders, they can provide exogenous variation in modifiable risk factors and thus serve as instruments.

Mendelian randomization has typically been employed to estimate the causal effects of factors such as body mass index (BMI), smoking, or alcohol consumption on outcomes including health status, educational attainment, and labour market performance. For instance, Ding, Lehrer, Rosenquist and Audrain-McGovern (2009) and Fletcher and Lehrer (2009) used genetic instruments to examine the impact of health conditions on academic achievement, while Norton and Han (2008) analysed the effects of obesity on labour market outcomes. In a similar vein, von Hinke Kessler Scholder, Smith, Lawlor, Propper and Windmeijer (2010) and von Hinke

Kessler Scholder, Smith, Lawlor, Propper and Windmeijer (2013) exploited genetic variation to study how children's physical characteristics influence human capital formation.

Challenges in the use of Mendelian randomization have been identified in the literature. In practice, the exclusion restriction assumption may be threatened by several biological and statistical mechanisms. Population stratification can generate spurious associations if genetic variation and outcomes differ systematically across subgroups. Pleiotropy, where a single genetic variant influences multiple traits, is particularly problematic if these traits independently affect the outcome. In addition, linkage disequilibrium may lead to the joint inheritance of variants that operate through distinct causal pathways. These issues have been widely discussed in both epidemiological and health economics research (Lawlor, Windmeijer & Davey Smith, 2008; Davey Smith, 2011). Genetic variants must also be sufficiently strongly associated with the phenotype of interest. This is typically supported by evidence from genome-wide association studies and replicated empirical findings across datasets (Davey Smith & Ebrahim, 2003). However, the strength of these associations may depend on environmental context due to gene–environment interactions. For example, Chen, Davey Smith, Harbord and Lewis (2008) have shown that genetic variants linked to alcohol metabolism have different effects on consumption behaviour depending on prevailing social norms regarding drinking, highlighting the contextual nature of genetic relevance. In addition, individual genetic variants typically explain only a small proportion of variation in complex traits leading to weak instruments. To address this limitation, many studies have constructed polygenic scores or allele indices that combine information across multiple genetic variants, thereby increasing explanatory power (Pierce, Ahsan & VanderWeele, 2011; Palmer et al., 2012). While this has improved statistical precision, it has raised additional concerns about the aggregation of variants with potentially heterogeneous or pleiotropic effects. Finally, Mendelian randomization analyses must also consider biological processes such as canalisation, whereby developmental mechanisms may compensate for genetic variation over time, attenuating the relationship between genotype and phenotype (Waddington, 1942). Although empirical evidence on the importance of canalisation in humans remains limited, it represents a conceptual limitation in interpreting genetic instrumental variable estimates and highlights the complexity of linking biological mechanisms to econometric identification strategies.

#### 18.5.1.2 From LATEs to Personalised Treatment Effects

An important conceptual advance in the IV literature was the recognition that such methods identify a local average treatment effect (LATE) (Imbens & Angrist, 1994). This insight has significant implications in health economics, where policy changes or institutional rules may only affect specific groups. The study of McClellan et al. (1994) illustrated that the benefit of intensive treatment for AMI for compliers may be less than for always-takers. Similarly, Cutler and Gruber (1996) used variation in Medicaid eligibility to estimate the effects of public insurance, focusing on individuals whose

coverage decisions responded to policy thresholds. Estimating the characteristics of compliers has become more common in recent IV studies. For example, von Hinke, Rice and Tominey (2022) demonstrate that the impact of maternal mental health problems during pregnancy on child outcomes are identified off vulnerable mothers with low education, an unintended pregnancy, or previous incidence of depression.

The evaluation literature, particularly that concerned with clinical outcomes research, has emphasised the role of essential heterogeneity in determining the treatment effects between alternative interventions that can inform personalised treatments. Recognising that individuals select into treatments based on their expected gains, and that such information is often unobserved, local instrumental variable (LIV) approaches have been developed to estimate mean treatment effect parameters in the presence of essential heterogeneity (Heckman & Vytlacil, 1999, 2005). In the context of applications to health economics, Basu, Heckman, Navarro-Lozano and Urzua (2007) illustrated its application to an analysis of breast cancer patients. The methods can be used to explore treatment effect heterogeneity across both observed and unobserved characteristics of individuals and can be used to estimate conditional average treatment effects (CATEs) based on observed factors. In the context of clinical evaluation, an important extension of the LIV approach was developed by Basu (2014) to estimate person-centred treatment (PeT) effects, which represent an individualised patient-specific treatment effect. While the standard LIV estimate is based on marginal treatment effects with respect to marginal changes in the propensity to treat, the PeT can be applied to specific individuals by aggregating MTE across the distribution of an individual's unobserved factors determining whether they receive treatment (resistance to treatment). Basu (2014) illustrated the use of PeT in an application to evaluating prostate cancer treatments.

#### 18.5.1.3 Nonlinear Models and IVs

The preponderance of limited-dependent variables, count variables, and skewed distributions in health economics, makes the application of IV techniques to nonlinear models vital. An early example is Mullahy (1997) who developed an IV approach for count data models in the presence of endogenous regressors estimated by generalized method of moments and applied to cigarette consumption. However, for the practical and general implementation of nonlinear IV approaches in health economics a notable contribution was provided by Terza, Basu and Rathouz (2008). The authors examined the properties of what were, at the time, two commonly applied approaches to the estimation of non-linear models with endogenous regressors: two-stage predictor substitution (2SPS) and two-stage residual inclusion (2SRI). The paper demonstrated that 2SRI estimation is generally consistent while 2SPS is not. The generalisability of the framework offered by 2SRI and its ease of implementation led to its wide application in health economics. For example, in investigating the role of competition and the impact of greater patient choice on the quality of care provided, Moscelli, Gravelle and Siciliani (2021) combined a 2SRI approach (with distance to providers as the instrument) to control for unobserved selection over provider choice within a

quasi-difference-in-differences framework. In the context of modelling healthcare costs in the presence of a binary treatment variable, Garrido, Deb, Burgess and Penrod (2012) considered 2SRI as a special case of the control function approach and specified, among other estimation methods, a generalised linear model using the Gamma family of distributions. Various residual types - response, Pearson, Anscombe, and deviance residuals - from the first-stage binary treatment model together with different polynomial functional forms of these were considered. Their results suggest caution in the a priori selection of models, particularly in cost regressions where incremental effects beyond those calculated at the mean are of interest.

#### 18.5.1.4 Partial Identification

Given the challenge of locating valid and relevant instruments, particularly within the context of individual health behaviours and the impact of health on broader social outcomes such as labour supply, it is perhaps surprising that methods based on partial identification have not received more attention in the health econometrics literature. The methods described by Altonji, Elder and Taber (2005) to assess the likely bias arising from unobservable variables - based on the selection effects from observable variables - are, in principle, well suited to the nonlinear models often encountered in health economics, and potentially informative by providing bounds on the treatment effect. Altonji, Elder and Taber (2008) illustrated the method by re-evaluating whether the Swan-Ganz catheterization used in intensive care raises or lowers mortality. Further applications in health econometrics have been sparse. Chatterji, Alegria and Takeuchi (2011) used the approach to investigate the impact of psychiatric disorders on employment and labour force participation by specifying a bivariate probit model. The sensitivity of the treatment effect to various values of the covariance in errors, $\rho$, was then assessed, including the value at which selection on unobservables is equal to selection on observables. Rice, Roberts and Sechel (2025) adapted the approach to a bivariate model with an ordered categorical outcome and continuous endogenous regressor to investigate the role of common mental health problems on a measure of work productivity.

Oster (2019) developed a similar approach for linear models and Bryan, Rice, Roberts and Sechel (2022) applied this to the estimation of panel data using a fixed effects linear probability model. Results show tight bounds on estimates of the impact of mental health on employment conditional on time-invariant individual heterogeneity. This suggests that unobserved selection into mental health problems is almost entirely based on time-invariant characteristics.

Similarly, partial identification in situations where the validity of an instrument is debatable (‘plausibly’ exogenous) set out in Conley, Hansen and Rossi (2012) has also received little attention. A novel application, however, was provided by Van Kippersluis and Rietveld (2018) who applied the approach to the use of genetic variants as IVs in the context of the impact of tobacco exposure on body mass index, and prostrate cancer on self-reported health. Acknowledging that the pleiotropic workings of genetic variants is not fully understood and that validity can be contested,

a zero-first stage to estimate the degree to which the IV exclusion restriction fails, was initially undertaken. This estimate was then used to condition the second-stage 2SLS estimation. The innovation in this application is through the use of a zero-first stage, which allows the parameter of the excluded instrument in the second stage to be estimated, rather than relying on assumptions about its distribution, as set out in Conley et al. (2012). Accordingly, a point estimate with its confidence interval, rather than only bounds, could be obtained.

### 18.5.2 Regression Discontinuity Designs

Regression discontinuity designs (RDD) provide another important quasi-experimental approach that has been used increasingly in health economics over recent years. The method exploits discontinuities in treatment assignment generated by threshold rules in policy or clinical decision-making and any associated discontinuity in outcome that arises. Individuals on either side of a cut-off are assumed to be comparable in all respects except for treatment status, allowing for the identification of local causal effects.

RDD is particularly well suited to health applications, where eligibility for treatment or insurance coverage often depends on observable thresholds in variables such as age, income, or clinical indicators. A canonical example is provided by Card, Dobkin and Maestas (2008), who exploited the age 65 eligibility threshold for Medicare in the United States to estimate the effects of insurance coverage on healthcare utilisation and health outcomes. D. S. Lee and Lemieux (2010) provided a comprehensive overview of regression discontinuity methods and highlight their applicability in health and other policy contexts.

A key advantage of RDD in health economics is its ability to address both observed and unobserved confounding in a local neighbourhood around the threshold. By focusing on individuals close to the threshold, the method effectively controls for unobserved heterogeneity that varies smoothly with the running variable. This makes RDD particularly credible in settings where selection into treatment is otherwise difficult to model. Unlike IV, RDD does not require an exclusion restriction; the running variable may be correlated with the outcome.

However, RDD estimates are inherently local to the threshold and may not generalise to broader populations, raising concerns about external validity. In addition, potential manipulation of the running variable—such as strategic reporting of income or clinical measures—can threaten identification. As a result, careful diagnostic checks and robustness analyses are essential components of applied RDD work.

### 18.5.3 Difference-In-Differences

Difference-in-differences (DID) methods exploit variation over time and across groups to identify causal effects, and in recent years have become one of the most widely used tools in applied health economics. The approach compares changes in outcomes for a treated group before and after an intervention with corresponding changes in a control group, relying on the key assumption of parallel trends. Under this assumption, in the absence of treatment, both groups would have evolved similarly over time. The method can be applied to longitudinal data or to repeated cross-sections.

DID has been extensively applied to evaluate health policy reforms and system-level interventions. For example, Cutler and Gruber (1996) study the effects of Medicaid expansions on insurance coverage, while Finkelstein (2007) analyse the impact of Medicare on healthcare spending and utilisation. These studies illustrate how DID methods allow researchers to exploit policy variation in real-world settings where randomization is not feasible.

In studying the effects of mandated employment-based insurance coverage for childbirth on labour market outcomes in the US, Gruber (1994) introduced a triple-difference estimator (DDD) as an extension of DID. Their concern was to control for systematic shocks to labour outcomes of treated individuals in experimental states which were correlated with the change to insurance laws. Using a DDD approach, changes in outcomes (e.g. wages) between the treated and controls in experimental states were compared to the same change in states where maternity mandates were not passed. In this setup the parallel trends assumption for identification is only required to hold for the difference in the ratios across the experimental and non-experimental states (see Olden & Møen, 2022). Gruber and Poterba (1994) further applied the DDD approach to study the effect of a health-insurance tax subsidy for the self-employed on the demand for insurance.

The DDD estimator has gained popularity particularly in applications where spillover effects from treated to controls within experimental settings is a concern (and hence a requirement for a non-experimental group), or where general economic conditions vary across experimental and non-experimental settings leading to differential pre-treatment trends. Nilsson (2017) applied a DDD approach to an aggregate-level analysis of labour market productivity of individuals exposed to alcohol in-utero. Exploiting a policy change that increased access to alcohol for young people in certain regions of Sweden, the design allowed them to control for region-specific shocks coincident with the policy which also affected child outcomes.

The availability of panel data has strengthened the application of DID in health economics, allowing researchers to control for individual fixed effects and time-varying confounders. This is particularly important given the persistence and state dependence observed in many health outcomes. Recent extensions of the basic DID framework allow for dynamic treatment effects, through event study specifications, and for heterogeneous responses to policies whose implementation is staggered over time (e.g., Goodman-Bacon, 2021).

Despite its widespread use, the validity of DID depends critically on the plausibility of the parallel trends assumption, which may not hold in many empirical settings.

Recent methodological developments, including synthetic control methods, have been introduced to assess and relax this assumption. In health economics applications, where policy changes are often accompanied by other concurrent reforms, careful specification, diagnostic testing, and robustness analysis remain essential for credible inference.

## 18.6 Discussion

As argued earlier, health economics has provided a rich field for the application of econometric techniques. The challenges created by health economic concepts and analysis along with the features of health data include latent variables, unobserved heterogeneity, endogeneity, nonlinear models, and the problems of handling survey and item non-response. These challenges have helped to stimulate methodological innovations and developments that have sometimes gone beyond health economics.

### 18.6.1 Some Neglected Topics

This chapter cannot claim to cover all aspects of the history of econometrics in health economics. In particular there are topics that have seen periods of intense activity and interest and then, for various reasons, have dropped off the agenda. This can be because of improvements in data availability, as well as changes in methodological focus. In particular, and in common with other areas of applied economics, there has been a shift in emphasis towards causal inference and the importance of credible identification strategies, accompanied by less reliance on methods that are sensitive to parametric assumptions and exclusion restrictions. Examples of such topics include the use of time series econometric methods to analyse aggregate data on health expenditures; stochastic frontier models and data envelopment analysis used to measure the efficiency of healthcare organisations; and conjoint analysis used to analyse the outcomes of discrete choice experiments.

### 18.6.2 Future Prospects

In recent years administrative datasets have superseded social surveys as the main source of data in applied health economics studies. These include healthcare provider reimbursement and claims databases, and population registers of births, deaths, cancer cases, etc. These datasets are collected for administrative purposes and may be made available to researchers. Administrative datasets often contain millions of observations and may cover a complete population, rather than just a random sample. As such they suffer from less unit and item non-response than survey data. They tend

to be less affected by reporting bias, but as they are collected routinely and on a wide scale they may be vulnerable to data input and coding errors. Administrative datasets are not designed by and for researchers, which means they may not contain all of the variables that would be of interest to researchers. Different data sources often have to be combined and linked administrative datasets are becoming more available.

Biological markers, or biomarkers, are becoming more prominent in research in health economics (Benzeval et al., 2016). Biomarkers are biological or physiological measures that indicate the presence of a disease or the propensity to develop a disease. They can be used to identify risk factors and as objective measures of health that avoid contamination by reporting bias. They have been incorporated into an increasing range of datasets, including longitudinal datasets, such as the US Health and Retirement Survey (HRS), the English Longitudinal Survey of Ageing (ELSA) and the UK Household Longitudinal Study (UKHLS). This has been enhanced by the availability of genetic and epigenetic data, which provide greater potential to control for individual heterogeneity.

The explosive growth of data science and artificial intelligence will inevitably shape the practice and content of research in health economics in the future. Machine learning approaches have been pioneered in the context of predictive modelling and improving the accuracy of predictions is of high relevance in the context of healthcare utilisation and expenditure. A particular area in which such methods have attracted attention is the literature on risk adjustment in US healthcare plans. Given the large number of Hierarchical Condition Category scores (HCCs), and disaggregated ICD-10 diagnosis codes available for risk adjustment, the use of machine learning is particularly appealing for optimising predictions of healthcare use (Rose, 2016). For example, in selecting subsets of HCCs from existing risk adjustment models without compromising statistical fit (McGuire, Zink & Rose, 2021), or to identify under-compensated groups of patients, where existing models fail to account fully for the additional costs from multiple conditions, or heterogeneity by condition with age and sex (Zink & Rose, 2021). Recently, the case has been made that supervised machine learning methods may be particularly useful in research for accurate and generalizable prediction of health outcomes (Padula et al., 2022). For example, Davillas and Jones (2025) have used the least absolute shrinkage and selection operator (LASSO) regression analysis to assess whether the availability of measures of epigenetic biological age enhances predictions of future GP consultations, outpatient visits and inpatient care.

The current literature in health economics has a greater focus on causal inference than on prediction and causal machine learning is coming to prominence. In particular, this encompasses methods for exploring heterogeneity in causal effects (Shah, Kreif & Jones, 2021). Causal machine learning provides a practical solution for high-dimensional settings where specifying parametric regressions can be challenging. Machine learning algorithms offer a flexible approach to subgroup analysis by selecting covariates of interest in a data-adaptive way. Machine learning also offers the prospect of going beyond causal inference to build data-driven policy rules for more efficient and equitable resource allocation. There is a growing literature on heterogeneous treatment effects and optimal policy learning in the context of

public health insurance. For example, Shah, Jones, Malenica, Hidayat and Kreif (2025) demonstrated how optimal policy learning might inform the targetting of Indonesia's two subsidised health insurance programmes. They used a 'super learner' ensembling approach, combining standard regression and machine learning algorithms to estimate conditional average treatment effects (CATE). These CATEs were then used to construct optimal policy rules.

## 18.7 Conclusions

In tracing the evolution of health econometrics, this chapter has highlighted how methodological developments have been shaped by the distinctive features of health and healthcare data, as well as by broader advances in econometric theory, and data availability. From the influence of the RAND Health Insurance Experiment to the increasing use of panel data, administrative records, and quasi-experimental methods, health economics has both adopted and contributed to innovations in empirical analysis. The field has adopted a rich toolkit for addressing challenges such as nonlinearity, endogeneity, unobserved heterogeneity, measurement error, and causal inference. This enables more rigorous investigation of health-related behaviours, outcomes, and policies.

The history of health econometrics also illustrates the dynamic relationship between research questions, data, and methodology. While some approaches have declined in prominence as new methods and data sources have emerged, many continue to inform contemporary research. Ongoing advances in data collection, linked administrative datasets, machine learning, and causal inference methods are likely to create new opportunities and challenges for researchers. As health systems face growing pressures associated with demographic change, technological innovation, and resource constraints, robust econometric analysis will remain central to generating evidence that informs effective and equitable health policy.

## References

Acton, J. P. (1975). Nonmonetary factors in the demand for medical services: some empirical evidence. *Journal of Political Economy*, *83*(3), 595–614.

Altonji, J., Elder, T. E. & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economics*, *113*(1), 151–184.

Altonji, J., Elder, T. E. & Taber, C. R. (2008). Using selection on observed variables to assess bias from unobservables when evaluating Swan-Ganz catheterization. *American Economic Review: Papers & Proceedings*, *98*(2), 345–350.

Álvarez, B. & Delgado, M. A. (2002). Goodness-of-fit techniques for count data models: an application to the demand for dental care in Spain. *Empirical*

*Economics*, *27*(3), 543–567.

Angelini, V., Cavapozzi, D. & Paccagnella, O. (2011). Dynamics of reporting work disability in Europe. *Journal of the Royal Statistical Society: Series A*, *174*(3), 621–638.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic Publishers.

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277–297.

Arellano, M. & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29–51.

Arinen, S., Sintonen, H. & Rosenqvist, G. (1996). *Dental utilisation by young adults before and after subsidisation reform in Finland*. University of York, Centre for Health Economics.

Aron-Dine, A., Einav, L. & Finkelstein, A. (2013). The RAND health insurance experiment, three decades later. *Journal of Economic Perspectives*, *27*(1), 197–222.

Arulampalam, W. & Bhalotra, S. (2006). Sibling death clustering in India: state dependence versus unobserved heterogeneity. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *169*(4), 829–848.

Atella, V., Belotti, F., Depalo, D. & Piano Mortari, A. (2014). Measuring spatial effects in the presence of institutional constraints: The case of Italian local health authority expenditure. *Regional Science and Urban Economics*, *49*, 232–241.

Auster, R., Leveson, I. & Sarachek, D. (1969). The production of health, an exploratory study. *Journal of Human Resources*, *4*(4), 411–436.

Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M. & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, *17*(3), 351–375.

Baicker, K. (2005). Spillover effects of state spending. *Journal of Public Economics*, *89*(2-3), 529–544.

Baker, M., Stabile, M. & Deri, C. (2004). What do self-reported, objective, measures of health measure? *Journal of Human Resources*, *39*(4), 1067–1093.

Baltagi, B., Moscone, F. & Santos, R. (2018). *Spatial Health Econometrics* (Vol. 127). Emerald Group Publishing Limited.

Baltagi, B. & Yen, Y. F. (2014). Hospital treatment rates and spillover effects: Does ownership matter? *Regional Science and Urban Economics*, *49*, 193–202.

Basu, A. (2014). Estimating person-centered treatment (pet) effects using instrumental variables: An applicaiton to evaluating prostate cancer treatments. *Journal of Applied Econometrics*, *29*(4), 671–691.

Basu, A., Heckman, J. J., Navarro-Lozano, S. & Urzua, S. (2007). Use of instrumental variablesin the presence of heterogensous and self-selection: An applicaiton to treatments of breast cancer. *Health Economics*, *16*(11), 1133–1157.

Basu, A. & Rathouz, P. J. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics*,

*6*(1), 93–109.

Bauer, T., Göhlmann, S. & Sinning, M. (2007). Gender differences in smoking behavior. *Health Economics*, *16*, 895–909.

Behrman, J. R., Sickles, R., Taubman, P. & Yazbeck, A. (1991). Black-white mortality inequalities. *Journal of Econometrics*, *50*(1-2), 183–203.

Behrman, J. R., Sickles, R. C. & Taubman, P. (1990). Age-specific death rates with tobacco smoking and occupational activity: Sensitivity to sample length, functional form, and unobserved frailty. *Demography*, *27*(2), 267–284.

Behrman, J. R. & Wolfe, B. L. (1987). How does mother's schooling affect family health, nutrition, medical care usage, and household sanitation? *Journal of Econometrics*, *36*(1-2), 185–204.

Benzeval, M., Kumari, M. & Jones, A. M. (2016). How do biomarkers and genetics contribute to understanding society? *Health Economics*, 1219–1222.

Blough, D. K., Madden, C. W. & Hornbrook, M. C. (1999). Modeling risk using generalized linear models. *Journal of Health Economics*, *18*(2), 153–171.

Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115–143.

Bolduc, D., Lacroix, G. & Muller, C. (1996). The choice of medical providers in rural Benin: a comparison of discrete choice models. *Journal of Health Economics*, *15*(4), 477–498.

Brookhart, M., Wang, P., Solomon, D. & Schneeweiss, S. (2006). Evaluating short-term drug effects using a physician-specifc prescribing preference as an instrumental variable. *Epidemiology*, *17*(3), 268–275.

Brown, T. T., Coffman, J. M., Quinn, B. C., Scheffler, R. M. & Schwalm, D. D. (2006). Do physicians always flee from HMOs? New results using dynamic panel estimation methods. *Health Services Research*, *41*(2), 357–373.

Bryan, M., Rice, N., Roberts, J. & Sechel, C. (2022). Mental health and employment: A bounding approach using panel data. *Oxford Bulletin of Economics and Statistics*, *84*(5), 1018–1051.

Buchmueller, T. C. & Feldstein, P. J. (1997). The effect of price on switching among health plans. *Journal of Health Economics*, *16*(2), 231–247.

Buntin, M. B. & Zaslavsky, A. M. (2004). Too much ado about two-part models and transformation?: Comparing methods of modeling Medicare expenditures. *Journal of Health Economics*, *23*(3), 525–542.

Butler, J. S., Anderson, K. H. & Burkhauser, R. V. (1989). Work and health after retirement: a competing risks model with semiparametric unobserved heterogeneity. *The Review of Economics and Statistics*, 46–53.

Börsch-Supan, A., Hajivassiliou, V. & Kotlikoff, L. J. (1992). Health, children, and elderly living arrangements: A multiperiod-multinomial probit model with unobserved heterogeneity and autocorrelated errors. In *Topics in the economics of aging* (pp. 79–108). University of Chicago Press.

Cameron, A. C. & Johansson, P. (1997). Count data regression using series expansions: with applications. *Journal of Applied Econometrics*, *12*(3), 203–223.

Cameron, A. C. & Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied*

*Econometrics*, *1*(1), 29–53.

Cameron, A. C., Trivedi, P. K., Milne, F. & Piggott, J. (1988). A microeconometric model of the demand for health care and health insurance in Australia. *The Review of Economic Studies*, *55*(1), 85–106.

Cameron, A. C. & Windmeijer, F. A. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics*, *14*(2), 209–220.

Card, D., Dobkin, C. & Maestas, N. (2008). The impact of nearly universal insurance coverage on health care utilization: evidence from medicare. *American Economic Review*, *98*(5), 2242–2258.

Carey, K. (2000). Hospital cost containment and length of stay: an econometric analysis. *Southern Economic Journal*, *67*(2), 363–380.

Cauley, S. D. (1987). The time price of medical care. *The Review of Economics and Statistics*, 59–66.

Chaloupka, F. & Wechsler, H. (1997). Price, tobacco control policies and smoking among young adults. *Journal of Health Economics*, *16*(3), 359–373.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, *47*(1), 225–238.

Chang, F.-R. & Trivedi, P. K. (2003). Economics of self-medication: theory and evidence. *Health Economics*, *12*(9), 721–739.

Chatterji, P., Alegria, M. & Takeuchi, D. (2011). Psychiatric disorders and labour market outcomes: Evidence from the National Comorbidity Survey - replication. *Journal of Health Economics*, *30*, 858—868.

Chen, L., Davey Smith, G., Harbord, R. M. & Lewis, S. J. (2008). Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Medicine*, *5*(3), e52.

Chernozhukov, V., Fernández-Val, I. & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205–2268.

Conley, T., Hansen, C. & Rossi, P. (2012). Plausibly exogenous. *The Review of Economics and Statistics*, *94*(1), 260–272.

Contoyannis, P., Jones, A. M. & Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*, *19*(4), 473–503.

Conway, K. S. & Deb, P. (2005). Is prenatal care really ineffective? Or, is the 'devil' in the distribution? *Journal of Health Economics*, *24*(3), 489–513.

Costa-Font, J. & Pons-Novell, J. (2007). Public health expenditure and spatial interactions in a decentralized national health system. *Health Economics*, *16*(3), 291–306.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, *39*(5), 829.

Cutler, D. M. & Gruber, J. (1996). The effect of Medicaid expansions on public insurance, private insurance, and redistribution. *The American Economic Review*, *86*(2), 378–383.

Davey Smith, G. (2011). Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. *Genes &*

*Nutrition*, *6*(1), 27–43.

Davey Smith, G. & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*(1), 1–22.

Davillas, A. & Jones, A. M. (2025). Biological age and predicting future health care utilisation. *Journal of Health Economics*, *99*, 102956.

Deaton, A. & Irish, M. (1984). Statistical models for zero expenditures in household budgets. *Journal of Public Economics*, *23*(1–2), 59–80.

Deb, P., Munkin, M. K. & Trivedi, P. K. (2006). Bayesian analysis of the two-part model with endogeneity: application to health care expenditure. *Journal of Applied Econometrics*, *21*(7), 1081–1099.

Deb, P. & Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics*, *12*(3), 313–336.

Deb, P. & Trivedi, P. K. (2002). The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics*, *21*(4), 601–625.

Deb, P. & Trivedi, P. K. (2006). Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization. *The Econometrics Journal*, *9*(2), 307–331.

Ding, W., Lehrer, S. F., Rosenquist, J. N. & Audrain-McGovern, J. (2009). The impact of poor health on academic performance: New evidence using genetic markers. *Journal of Health Economics*, *28*(3), 578–597.

Disney, R., Emmerson, C. & Wakefield, M. (2006). Ill health and retirement in Britain: A panel data-based analysis. *Journal of Health Economics*, *25*(4), 621–649.

Donaldson, C., Jones, A. M., Mapp, T. J. & Olson, J. A. (1998). Limited dependent variables in willingness to pay studies: applications in health care. *Applied Economics*, *30*(5), 667–677.

Douglas, S. (1998). The duration of the smoking habit. *Economic Inquiry*, *36*(1), 49–64.

Douglas, S. & Hariharan, G. (1994). The hazard of starting smoking: estimates from a split population duration model. *Journal of Health Economics*, *13*(2), 213–230.

Dowd, B., Feldman, R., Cassou, S. & Finch, M. (1991). Health plan choice and the utilization of health care services. *The Review of Economics and Statistics*, 85–93.

Doyle Jr, J. J. (2005). Health insurance, treatment and outcomes: using auto accidents as health shocks. *Review of Economics and Statistics*, *87*(2), 256–270.

Dranove, D. & Wehner, P. (1994). Physician-induced demand for childbirths. *Journal of Health Economics*, *13*, 61–73.

Duan, N., Manning, W. G., Morris, C. N. & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, *1*(2), 115–126.

Enami, K. & Mullahy, J. (2009). Tobit at fifty: a brief history of Tobin's remarkable estimator, of related empirical methods, and of limited dependent variable econometrics in health economics. *Health Economics*, *18*(6), 619–628.

Evans, W. N., Farrelly, M. C. & Montgomery, E. (1999). Do workplace smoking bans reduce smoking? *American Economic Review*, *89*(4), 728–747.

Feldman, R., Finch, M., Dowd, B. & Cassou, S. (1989). The demand for employment-based health insurance plans. *Journal of Human Resources*, 115–142.

Feldstein, M. S. (1967). *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service* (Vol. 51). Amsterdam: North-Holland Publishing Company.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of Medicare. *The Quarterly Journal of Economics*, *122*(1), 1–37.

Fletcher, J. M. & Lehrer, S. F. (2009). Using genetic lotteries within families to examine the causal impact of poor health on academic achievement. *National Bureau of Economic Research WP*, *15148*.

Foresi, S. & Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, *90*(430), 451–466.

French, M. T. & Popovici, I. (2011). That instrument is lousy! In search of agreement when using instrumental variables estimation in substance use research. *Health Economics*, *20*, 127–146.

Gannon, B. (2005). A dynamic analysis of disability and labour force participation in Ireland 1995–2000. *Health Economics*, *14*(9), 925–938.

Garrido, M., Deb, P., Burgess, J. F. & Penrod, J. D. (2012). Choosing models for health care cost analyses: Issues of nonlinearity and endogeneity. *Health Sevices Research*, *46*(6), 2377–2397.

Gerdtham, U.-G. (1997). Equity in health care utilization: further tests based on hurdle models and Swedish micro data. *Health Economics*, *6*(3), 303–319.

Gertler, P., Locay, L. & Sanderson, W. (1987). Are user fees regressive?: The welfare implications of health care financing proposals in Peru. *Journal of Econometrics*, *36*(1-2), 67–88.

Gilleskie, D. B. & Mroz, T. A. (2004). A flexible approach for estimating the effects of covariates on health expenditures. *Journal of Health Economics*, *23*(2), 391–418.

Godøy, A., Haaland, V. F., Huitfeldt, I. & Votruba, M. (2024). Hospital queues, patient health and labor supply. *American Economic Journal: Economic Policy*, *16*(2), 150–181.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Gravelle, H., Santos, R. & Siciliani, L. (2014). *Does a hospital's quality depend on the quality of other hospitals? A spatial econometrics approach* (Vol. 49).

Greene, W., Harris, M., Knott, R. & Rice, N. (2021). Specification and testing of hierarchical ordered response models with anchoring vignettes. *Journal of the Royal Statistical Society: Series A*, *184*, 31–64.

Grol-Prokopczyk, H., Freese, J. & Hauser, R. (2011). Using anchoring vignettes to assess group differences in general self-rated health. *Journal of Health and Social Behavior*, *52*(2), 246–261.

Grootendorst, P. V. (1997). Health care policy evaluation using longitudinal insurance

claims data: an application of the panel tobit estimator. *Health Economics*, *6*(4), 365–382.

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, *84*(3), 622–41.

Gruber, J. & Poterba, J. (1994). Tax incentives and the decision to purchase health insurance.: Evidence from the self-employed. *Quarterly Journal of Economics*(3), 701–33.

Gurmu, S. (1997). Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. *Journal of Applied Econometrics*, *12*(3), 225–242.

Haas-Wilson, D. & Savoca, E. (1990). Quality and provider choice: A multinomial logit-least-squares model with selectivity. *Health Services Research*, *24*(6), 791.

Hamilton, B. H. (1999). HMO selection and Medicare costs: Bayesian MCMC estimation of a robust panel data tobit model with survival. *Health Economics*, *8*(5), 403–414.

Hamilton, B. H. & Hamilton, V. H. (1997). Estimating surgical volume—outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Economics*, *6*(4), 383–395.

Han, A. & Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, *5*(1), 1–28.

Hauck, K. & Rice, N. (2004). A longitudinal analysis of mental health mobility in Britain. *Health Economics*, *13*(10), 981–1001.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, *49*, 1251-1271.

Hay, J. W. & Olsen, R. J. (1984). Let them eat cake: a note on comparing alternative models of the demand for medical care. *Journal of Business & Economic Statistics*, *2*(3), 279–282.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 153–161.

Heckman, J. J. & Vytlacil, E. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences USA*, *96*(8), 4730–4734.

Heckman, J. J. & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, *73*, 669–738.

Ho, V., Hamilton, B. H. & Roos, L. L. (2000). Multiple approaches to assessing the effects of delays for hip fracture patients in the United States and Canada. *Health Services Research*, *34*, 1499–1518.

Hoe, T. P. (2022, May). Does hospital crowding matter? Evidence from trauma and orthopedics in england. *American Economic Journal: Economic Policy*, *14*(2), 231-62.

Hoerger, T. J., Picone, G. A. & Sloan, F. A. (1996). Public subsidies, private provision of care and living arrangements of the elderly. *The Review of Economics and Statistics*, 428–440.

Holly, A. (2009). *Modelling risk using fourth order pseudo maximum likelihood*

*methods* (Tech. Rep.). Lausanne, Switzerland: Institute of Health Economics and Management (IEMS), University of Lausanne.

Idler, E. L. & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, 21–37.

Imbens, G. W. & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Jones, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics*, *4*, 23–39.

Jones, A. M. (2000). Health econometrics. In *Handbook of Health Economics* (Vol. 1, pp. 265–344). Elsevier.

Jones, A. M. (2009). Panel data methods and applications to health economics. *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics*, 557–631.

Jones, A. M. (2012). Models for health care. In M. P. Clements & D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting* (pp. 625–654). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780195398649.013.0024

Jones, A. M. & Labeaga, J. M. (2003). Individual heterogeneity and censoring in panel data estimates of tobacco expenditure. *Journal of Applied Econometrics*, *18*(2), 157–177.

Jones, A. M., Lomas, J. & Rice, N. (2014). Applying beta-type size distributions to healthcare cost regressions. *Journal of Applied Econometrics*, *29*(4), 649–670.

Jones, A. M., Lomas, J. & Rice, N. (2015). Healthcare cost regressions: going beyond the mean to estimate the full distribution. *Health Economics*, *24*(9), 1192–1212.

Jones, A. M. & Rice, N. (2012). Econometric evaluation of health policies. In *The Oxford Handbook of Health Economics.* Oxford University Press.

Jones, A. M., Rice, N., Smith, P., Ginnelly, L. & Sculpher, M. (2005). Using longitudinal data to investigate socioeconomic inequality in health. In *Health policy and economics: opportunities and challenges* (pp. 88–120). Open University Press.

Kapteyn, A., Smith, J. & Van Soest, A. (2007). Vignettes and self-reports of work disability in the United States and the Netherlands. *The American Economic Review*, *97*, 461–473.

Kenkel, D. S. (1995). Should you eat breakfast? Estimates from health production functions. *Health Economics*, *4*(1), 15–29.

Kerkhofs, M. & Lindeboom, M. (1995). Subjective health measures and state dependent reporting errors. *Health Economics*, *4*(3), 221–235.

King, G., Murray, C., Salomon, J. & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–207.

Kreider, B. (1999). Latent work disability and reporting bias. *Journal of Human Resources*, *34*, 734-769.

Lawlor, D. A., Windmeijer, F. & Davey Smith, G. (2008). Is Mendelian randomization 'lost in translation?': Comments on 'Mendelian randomization equals instru-

mental variable analysis with genetic instruments' by Wehby et al. *Statistics in Medicine*, *27*(15), 2750–2755.

Lee, D. S. & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355.

Lee, L.-F., Rosenzweig, M. R. & Pitt, M. M. (1997). The effects of improved nutrition, sanitation, and water quality on child health in high-mortality populations. *Journal of Econometrics*, *77*(1), 209–235.

Lee, M.-C. & Jones, A. M. (2004). How did dentists respond to the introduction of global budgets in Taiwan? An evaluation using individual panel data. *International Journal of Health Care Finance and Economics*, *4*(4), 307–326.

Leung, S. F. & Yu, S. (1996). On the choice between sample selection and two-part models. *Journal of Econometrics*, *72*(1-2), 197–229.

Levinson, A. & Ullman, F. (1998). Medicaid managed care and infant health. *Journal of Health Economics*, *17*(3), 351–368.

Lindeboom, M., Portrait, F. & Van den Berg, G. J. (2002). An econometric analysis of the mental-health effects of major events in the life of older individuals. *Health Economics*, *11*(6), 505–520.

Longo, F., Siciliani, L., Gravelle, H. & Santos, R. (2017). Do hospitals respond to rivals' quality and efficiency?: A spatial econometrics approach. *Health Economics*, *26*, 38–62.

Maddala, G. S. (1985). A survey of the literature on selectivity bias as it pertains to health care markets. *Advances in Health Economics and Health Services Research*, *6*, 3–26.

Mainous, A. G. & Gill, J. M. (1998). The importance of continuity of care in the likelihood of future hospitalization: is site of care equivalent to a primary clinician? *American Journal of Public Health*, *88*(10), 1539–1541.

Manning, W. G. (1998). The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics*, *17*(3), 283–295.

Manning, W. G. (2006). Dealing with skewed data on costs and expenditures. In A. M. Jones (Ed.), *The Elgar Companion to Health Economics* (pp. 439–446). Cheltenham, UK: Edward Elgar Publishing.

Manning, W. G., Basu, A. & Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, *24*(3), 465–488.

Manning, W. G. & Mullahy, J. (2001). Estimating log models: to transform or not to transform? *Journal of Health Economics*, *20*(4), 461–494.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B. & Leibowitz, A. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *The American Economic Review*, 251–277.

Manski, C. F. (1993). The selection problem in econometrics and statistics. *Handbook of Statistics*, *11*, 73-84.

McClellan, M., McNeil, B. J. & Newhouse, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: Analysis using instrumental variables. *Journal of the American Medical Association*, *272*(11), 859–866.

McGuire, T., Zink, G. & Rose, S. (2021). Improving the performance of risk adjustment systems: Constrained regressions, reinsurance and variable selection. *American Journal of Health Economics*, *7*(4), 497–521.

Mobley, L. R. (2003). Estimating hospital market pricing: An equilibrium approach using spatial econometrics. *Regional Science and Urban Economics*, *33*(4), 489–516.

Morris, C. N., Norton, E. C. & Zhou, X. H. (1994). Parametric duration analysis of nursing home usage. *Case Studies in Biometry*, 231–248.

Moscelli, G., Gravelle, H. & Siciliani, L. (2021). Hospital competition and quality for non-emergency patients in the English NHS. *Rand Journal of Economics*, *52*(2), 382-–414.

Moscone, F., Knapp, M. & Tosetti, E. (2007). Mental health expenditure in England: a spatial panel approach. *Journal of Health Economics*, *26*(4), 842–864.

Moscone, F. & Tosetti, E. (2014). *Spatial econometrics: Theory and applications in health economics.* Elsevier+.

Moscone, F., Tosetti, E. & Vittadini, G. (2011). Social interaction in patients' hospital choice: Evidence from Italy. *Journal of the Royal Statistical Society: Series A*, 453–472.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, *33*, 341–365.

Mullahy, J. (1997). Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics*, *79*(4), 586–593.

Mullahy, J. (1998). Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics*, 247–281.

Mullahy, J. & Norton, E. (2023). Why transform Y? The pitfalls of transformed regressions with a mass at zero. *Oxford Bulletin of Economics and Statistics*, *86*(2), 417–447.

Mundlak, Y. (1978). On the pooling of time series and cross-sectional data. *Econometrica*, *46*(1), 69–85.

Munkin, M. K. & Trivedi, P. K. (1999). Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. *The Econometrics Journal*, *2*(1), 29–48.

Neelon, B., Ghosh, P. & Loebs, P. F. (2011). A spatial poisson hurdle model for exploring geographic variation in emergency department visits. *Journal of the Royal Statistical Society: Series A*, *176*(2), 389–413.

Newhouse, J. P. (1987). Health economics and econometrics. *The American Economic Review*, *77*(2), 269–274.

Newhouse, J. P., Phelps, C. E. & Marquis, M. S. (1980). On having your cake and eating it too: Econometric problems in estimating the demand for health services. *Journal of Econometrics*, *13*(3), 365–390.

Nilsson, J. (2017). Alcohol availability, prenatal conditions, and long-term economic outcomes. *Journal of Political Economy*, *125*(4), 1149–1207.

Nolan, A. (2007). A dynamic analysis of GP visiting in Ireland: 1995–2001. *Health Economics*, *16*(2), 129–143.

Norton, E. C. (1995). Elderly assets, Medicaid policy, and spend-down in nursing homes. *Review of Income and Wealth*, *41*(3), 309–329.

Norton, E. C. & Han, E. (2008). Genetic information, obesity, and labor market outcomes. *Health Economics*, *17*(9), 1089–1104.

Olden, A. & Møen, J. (2022). The triple difference estimator. *Econometrics Journal*, *25*, 531–553.

Ortiz, L. G. G. & Masiero, G. (2013). Disentangling spillover effects of antibiotic consumption: A spatial panel approach. *Applied Economics*, *45*(8), 1041–1054.

Oster, E. (2019). Unobservable selection and coefficient stability: theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-–204.

Paccagnella, O. (2011). Anchoring vignettes with sample selection due to non-response. *Journal of the Royal Statistical Society: Series A*, *174*(3), 665–687.

Padula, W. V., Kreif, N., Vanness, D. J., Adamson, B., Rueda, J.-D., Felizzi, F., ... Crown, W. (2022). Machine learning methods in health economics and outcomes research—the PALISADE checklist: a good practices report of an ISPOR task force. *Value in Health*, *25*(7), 1063–1080.

Palmer, T. M., Lawlor, D. A., Harbord, R. M., Sheehan, N. A., Tobias, J. H., Timpson, N. J., ... Sterne, J. A. (2012). Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*, *21*(3), 223–242.

Peracchi, F. & Rossetti, C. (2012). Heterogeneity in health responses and anchoring vignettes. *Empirical Economics*, *42*(2), 513–538.

Peracchi, F. & Rossetti, C. (2013). The heterogeneous thresholds ordered response model: identification and inference. *Journal of the Royal Statistical Society, Series A*, *176*(3), 703–722.

Philipson, T. (1996). Private vaccination and public health: an empirical examination for US measles. *Journal of Human Resources*, 611–630.

Pierce, B. L., Ahsan, H. & VanderWeele, T. J. (2011). Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *International Journal of Epidemiology*, *40*(3), 740–752.

Pitt, M. M. (1997). Estimating the determinants of child health when fertility and mortality are selective. *Journal of Human Resources*, 129–158.

Pohlmeier, W. & Ulrich, V. (1995). An econometric model of the two-part decision-making process in the demand for health care. *Journal of Human Resources*, 339–361.

Pudney, S. & Shields, M. (2000). Gender, race, pay and promotion in the British nursing profession: estimation of a generalized ordered probit model. *Journal of Applied Econometrics*, *15*(4), 367-399.

Rashad, I. & Kaestner, R. (2004). Teenage sex, drugs, and aclcohol use: problems identifying the cause of risky behaviors. *Journal of Health Economics*, *23*, 493–503.

Rice, N., Roberts, J. & Sechel, C. (2025). Mental health and labout productivity. *Journal of Economic Behavior and Organization*, *236*.

Rice, N., Robone, S. & Smith, P. (2012). Vignettes and health systems responsiveness in cross-country comparative analyses. *Journal of the Royal Statistical Society, Series A*, *175*(2), 337-–369.

Riphahn, R. T., Wambach, A. & Million, A. (2003). Incentive effects in the demand for health care: a bivariate panel count data estimation. *Journal of Applied Econometrics*, *18*(4), 387–405.

Rose, S. (2016). A machine learning framework for plan payment risk adjustment. *Health Services Research*, *51*(6), 2358–2374.

Rosenzweig, M. R. & Schultz, T. P. (1983). Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight. *Journal of Political Economy*, *91*(5), 723–746.

Rosett, R. N. & Huang, L.-f. (1973). The effect of health insurance on the demand for medical care. *Journal of Political Economy*, *81*(2, Part 1), 281–305.

Santos Silva, J. & Windmeijer, F. (2001). Two-part multiple spell models for health care demand. *Journal of Econometrics*, *104*, 67-89.

Sarma, S. & Simpson, W. (2006). A microeconometric analysis of Canadian health care utilization. *Health Economics*, *15*(3), 219–239.

Shah, V., Jones, A. M., Malenica, I., Hidayat, T. & Kreif, N. (2025). Using policy learning to inform health insurance targeting: a case study of Indonesia. *Health Economics*, *34*(12), 2270–2296.

Shah, V., Kreif, N. & Jones, A. M. (2021). Machine learning for causal inference: estimating heterogeneous treatment effects. In N. Hashimzade & M. A. Thornton (Eds.), *Handbook of Research Methods and Applications in Empirical Microeconomics.* Edward Elgar Publishing.

Sheu, M.-L., Hu, T.-W., Keeler, T., Ong, M. & Sung, H.-Y. (2004). The effect of major cigarette price change on smoking behavior in california: a zero-inflated negative binomial model. *Health Economics*, *13*, 721-791.

Shorrocks, A. (1978). Income inequality and income mobility. *Journal of Economic Theory*, *19*(2), 376–393.

Sirven, N., Santos-Eggimann, B. & Spagnoli, J. (2012). Comparability of health care responsiveness in Europe. *Social Indicators Research*, *105*(2), 255–271.

Soloman, J., Tandon, A. & Murray, C. (2004). Comparability of selfrated health: cross-sectional multi-country survey using anchoring vignettes. *British Medical Journal*, *328*, 258–260.

Stern, S. (1996). Semiparametric estimates of the supply and demand effects of disability on labor force participation. *Journal of Econometrics*, *71*(1-2), 49–70.

Stoye, G. & Zaranko, B. (2020). *How accurate are self-reported diagnoses? Comparing self-reported health events in the English Longitudinal Study of Ageing with administrative hospital records.* Institute for Fiscal Studies.

Street, A., Jones, A. M. & Furuta, A. (1999). Cost sharing and pharmaceutical utilisation and expenditure in russia. *Journal of Health Economics*, *18*(4), 459–472.

Sutton, M. & Godfrey, C. (1995). A grouped data regression approach to estimating economic and social influences on individual drinking behaviour. *Health*

*Economics*, *4*(3), 237–247.

Tamm, M., Tauchmann, H., Wasem, J. & Greß, S. (2007). Elasticities of market shares and social health insurance choice in Germany: a dynamic panel data approach. *Health Economics*, *16*(3), 243–256.

Terza, J. V., Basu, A. & Rathouz, P. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, *27*, 531–543.

Theodossiou, I. (1998). The effects of low-pay and unemployment on psychological well-being: A logistic regression approach. *Journal of Health Economics*, *17*(1), 85–104.

Urban, N., Anderson, G. L. & Peacock, S. (1994). Mammography screening: how important is cost as a barrier to use? *American Journal of Public Health*, *84*(1), 50–55.

Van de Ven, W. P. & Van der Gaag, J. (1982). Health as an unobservable; a MIMIC model for health care demand. *Journal of Health Economics*, *1*, 157–183.

Van de Ven, W. P. & Van Praag, B. M. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics*, *17*(2), 229–252.

Van Doorslaer, E. & Koolman, X. (2004). Explaining the differences in income-related health inequalities across European countries. *Health Economics*, *13*(7), 609–628.

Van Kippersluis, H. & Rietveld, C. (2018). Pleiotropy-robust mendelian randomization. *International Journal of Epidemiology*, *47*(4), 1279–1288.

Vanness, D. J. & Mullahy, J. (2006). Perspectives on mean-based evaluation of health care. In A. M. Jones (Ed.), *The Elgar Companion to Health Economics* (chap. 49). Cheltenham, UK: Edward Elgar Publishing.

Van Ourti, T. (2004). Measuring horizontal inequity in Belgian health care using a Gaussian random effects two part count data model. *Health Economics*, *13*(7), 705–724.

Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A. & Smith, J. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society, Series A*, *174*(3), 575—595.

Van Vliet, R. J. & Van Praag, B. M. (1987). Health status estimation on the basis of MIMIC-health care models. *Journal of Health Economics*, *6*(1), 27–42.

Vistnes, J. P. & Hamilton, V. (1995). The time and monetary costs of outpatient care for children. *The American Economic Review*, *85*(2), 117–121.

von Hinke, S., Rice, N. & Tominey, E. (2022). Mental health around pregnancy and child development from early childhood to adolescence. *Labour Economics*, *78*.

von Hinke Kessler Scholder, S., Smith, G. D., Lawlor, D. A., Propper, C. & Windmeijer, F. (2010). Genetic markers as instrumental variables: An application to child fat mass and academic achievement. *Cemmap Working Paper*, *10*(229).

von Hinke Kessler Scholder, S., Smith, G. D., Lawlor, D. A., Propper, C. & Windmeijer, F. (2011). Mendelian randomization: the use of genes in instrumental variable

analyses. *Health Economics*, *20*(8), 893–896.

von Hinke Kessler Scholder, S., Smith, G. D., Lawlor, D. A., Propper, C. & Windmeijer, F. (2013). Child height, health and human capital: evidence using genetic markers. *European Economic Review*, *57*, 1–22.

Vonkova, H. & Hullegie, P. (2011). Is the anchoring vignettes method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society, Series A*, *174*(3), 597–620.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, *150*(3811), 563–565.

Wagstaff, A. (1986). The demand for health: some new empirical evidence. *Journal of Health Economics*, *5*(3), 195–233.

Wagstaff, A. (1989). Econometric studies in health economics: A survey of the British literature. *Journal of Health Economics*, *8*(1), 1–51.

Wagstaff, A., Van Doorslaer, E. & Paci, P. (1989). Equity in the finance and delivery of health care: some tentative cross-country comparisons. *Oxford Review of Economic Policy*, *5*(1), 89–112.

Wagstaff, A., Van Doorslaer, E. & Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics*, *112*(1), 207–223.

Winkelmann, R. (2004). Health care reform and the number of doctor visits—an econometric analysis. *Journal of Applied Econometrics*, *19*(4), 455–472.

Wolfe, B. & van der Gaag, J. (1981). A new health status index for children. In J. van der Gaag & M. Perlman (Eds.), *Health, Economics, and Health Economics.* Amsterdam: North-Holland.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics*, *20*(1), 39–54.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.

Wooldridge, J. M. (2019). Correlated random effects model swith unbalanced panels. *Journal of Econometrics*, *211*(1), 137–150.

Yen, S. T., Tang, C.-H. & Su, S.-J. (2001). Demand for traditional medicine in Taiwan: A mixed Gaussian-Poisson model approach. *Journal of Econometrics*, *10*, 221–232.

Zimmer, D. M. & Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, *24*(1), 63–76.

Zink, A. & Rose, S. (2021). Identifying undercompensated groups defined by multiple attributes in risk adjustment. *BMJ Health & Care Informatics*, *28*.