EVALUATING THE IMPACT OF HEALTHCARE POLICIES IN HUNGARY USING INDIVIDUAL-LEVEL DATA

A practical guide to estimation methods[1,2]

Péter Elek[3]

1. Introduction

This practical guide gives an overview of the possibilities of using individual level administrative data in the ex post impact evaluation of health policies and of implementing such impact evaluations in Hungary. The guide is intended to support the monitoring of health policy reforms between 2006 and 2008, but may also prove useful for other impact evaluations.

Ex post evaluations aim to distinguish between the actual impact of and intervention from the impact of other, simultaneous factors. This can be carried out more precisely on the basis of individual rather than aggregate data; moreover, individual data can help identify certain relationships (e.g. higher exposure of certain groups) that aggregate data cannot.

A good overview of the methodological issues of impact evaluations with a focus on health policy is provided by Khandker et al. (2010) and Grun (2006). For a more thorough understanding of econometric techniques the textbooks by Woolridge (2002 and 2009) can provide useful help.

2. Prerequisites for impact evaluations

A prerequisite for impact evaluations is that the characteristics of the intervention are clearly defined: when it happened, who was and who was not affected, when (theoretically) could it have an impact and on what outcomes. More precisely, it is necessary that those affected by the intervention can be identified either on the basis of time or other criteria, and the expected outcomes are measurable. Finally, a large enough sample of data on treatment and non-treatment groups and outcome variables should be available for the analysis.

3. Data

Data sources. The main sources of data for impact evaluation are the health insurance databases of individual cases that include for each case

[1] The author is grateful for useful comments by Ágota Scharle on earlier versions of this note.

[2] Translated by Ágnes Kozma Turnpenny and edited by Ágota Scharle.

[3] Economics Department, Eötvös Loránd University, Budapest, Hungary

- an identifier that uniquely identifies individuals in the data (but cannot be linked to personal records)

- the outcome of the treatment (result, duration of the treatment etc.),

- the medical variables that influence the outcome (diagnoses, interventions, hospital days, co-morbidity, medication etc.),

- the location of treatment (town, institution, hospital unit) and the related funding and other characteristics of the institution, hospital unit (e.g. output volume cap, number of beds),

- the main demographic characteristics of the individual (age, place of residence etc.) and the related socio-economic indicators of the local area (e.g. type of settlement, unemployment rate etc.). Though technically more difficult, it is possible to link health care records with data from other administrative (e.g. social insurance) databases at the individual level (e.g. the receipt of welfare payments, employment-related data etc.), which provide additional information on health status.

The structure of the data. Micro-level data from the above sources is usually pooled cross-sectional data, which means that data are collected at different points of time and from different samples. (However, some individuals may happen to be observed multiple times in a dataset.) If the same people are observed at different points in time, it is referred to as panel data, and if the data are taken at the same single time point it is a simple cross-sectional sample.

4. The steps of an impact evaluation

4.1. Defining the intervention and the expected impact, types of variables

Intervention. As the first step of the impact evaluation, the "treatment" (intervention) and the expected impact should be clearly defined. If it can be established that certain individuals received the "treatment" at certain times, and others / at other times did not receive it, then the intervention can be described with a dummy variable with values of 0 and 1. (For example, for the visit fee introduced in Hungary in February 2007, the value of the variable for individuals obliged to pay the fee in the given time period is 1, all else 0.) If the intervention directly affects a continuous variable then this should be used as the variable. For example for the reduction of the number of hospital beds a dummy variable (i.e. whether the number of hospital beds decreased or not) is not the best option. Instead, the changes expressed as a percentage of the number of beds over the previous year (or six months) give a more accurate measure of the effectiveness of the intervention. Similarly, all variables in the database can be continuous or discrete.

Expected impact. The expected impact of the intervention is measured in terms of the improvement or deterioration of an indicator (outcome variable). The outcome variable can also be continuous or discrete (in this case it is usually a dummy with a value of 0 or 1). A variant of the continuous outcome variable is the duration variable which measures the time until a specific event occurred (death, re-admission to hospital etc.). For example it might be relevant when people with mental health problems who have been discharged from hospital are readmitted. In this case the outcome variable is a duration variable, however it can also be

turned into a dummy for example by looking at the percentage of patients re-admitted to the hospital within 30 days of discharge and thus capturing the medically relevant information in a 0-1 variable. (The value of the variable is 1 if the participant has been re-admitted within 30 days, and 0 otherwise.)

The outcome variable (or some of the explanatory variables) may be censored, i.e. for some cases the data show only that it was above or below a certain level, but its exact value is unobserved. This is often the case with duration variables, e.g. for patients discharged from hospital but not yet readmitted, we only know that their readmission time is greater than the observed time elapsed since they were discharged. This should be taken into account when choosing the appropriate estimation method.

4.2 Introduction to outcome and explanatory variables, descriptive statistics, data cleaning

Univariate descriptive statistical analysis. Before the econometric analysis of the data, it is useful to analyse the data using descriptive statistics (mean, variance, quartiles, minimum and maximum values, outlier values, and a histogram). By looking at outliers possible typos can be picked up and corrected. Some further cleaning of the data might also be necessary to ensure internal logical consistency (to filter out contradictions). Missing data should be analysed separately (which variables are affected, how many cases are missing and why) because regressions will be estimated using the narrowest subsample (i.e. where there are no missing variables). Often missing data refers to 0 values which should be added. If the missing data are the logarithm of another variable with 0 values (i.e. the "value" of the transformed variable is $-\infty$), it might be useful to enter a negative number with a very large absolute value before running the regressions.

Multivariate desriptive statistics. As the next step of the descriptive statistical analysis, the raw (without controlling for other variables) relationship between the outcome and intervention variables should be analysed. If the intervention variable is a dummy (0-1 values) then the subgroup means of the outcome variable can be calculated for both values. The same subgroup means can also be calculated along other variables (year, age group etc.). For example people under 18 years did not have to pay the visit fee that was in force between February 2007 and 2008, therefore the value of the "Visit fee" intervention variable is 1 only for adults and during this period, otherwise it is 0. As an initial descriptive statistic it is worthwhile to calculate the mean value of the outcome variable for adults and children for each month. A graphical representation of the values would reveal if there were any changes in the difference between the mean values of the outcome variable for adults and children in 2007 compared to other years, which might (initially) indicate the impact of the visit fee. If the intervention variable is a continuous variable, subgroup means are usually calculated for high, medium and low values.

4.3 Identifying the impact

After defining the variables, the preliminary analysis of the data shows how the impact of the interventions can be identified. In some cases the impact of an intervention can only be captured in the variation of a variable across time (e.g. the introduction of public waiting lists), in other cases there may be cross-sectional variation, i.e. across groups that are more or less affected by the intervention. An example of this is the visit fee already mentioned above, or if the reduction of hospital beds was different across counties this would bring some additional

variance into the intervention variable in addition to the time dimension.

Regressional framework for identifying the impact. In ideal cases (but health policy impact evaluations are seldom ideal cases), "treatment" and control goups (those not affected by the intervention) are identical in terms of all the other characteristics. (This is the case in designed experiments in natural sciences.) In this case the impact of the intervention can be easily estimated from the difference between the outcome of the treatment and control groups. In practice however, the difference that can be attributed to the difference in the characteristics of the two groups should be separated from the impact of the intervention. Thus, the estimation method should control for other factors that may have had a differential effect on the treatment and the control group. In a regression analysis, explanatory variables can be added to control for these outside effects.

It is important to note that parameters estimated in a regression (including that for an intervention variable) may only be interpreted as measuring a causal relationship (effect) if explanatory variables are *exogenous*, i.e. all other relevant factors are controlled for. This is often difficult and exogeneity may be violated (endogeneity may arise) for several non-obvious reasons, some of which (unobservable variables, selection bias) are discussed in section 5 below.

The problem of endogeneity is easier to tackle by using instrumental variables (in cross section data) or using panel data than in a simple cross sectional regression (see e.g. Wooldridge 2002 or 2009). In panel data, unobserved individual characteristics can be explicitely controlled for. There are several other methods for identifying the impact of an intervention, for example discontinuity models and matching methods that are not discussed in detail here (see e.g. Khandker et el. 2010).

Transmission mechanism. The impact of an intervention on the outcome variable is often indirect, so that intermittent factors (the link between the intervention and the outcome) must be identified. The transmission mechanism describes the process whereby the intervention affects the outcome (e.g. the redrawing of GP districts or hospital catchment areas would affect health outcomes via disrupting doctor-patient relationships). These mechanisms may be identified by consulting healthcare professionals. When measuring the intermittent factors, one may need to find proxy variables, i.e. a descrete or continuous variable that adequately captures the transmission mechanism (in the above example, the volatility of patient-doctor relationships max be indicated by the rate of patients redirected to another institution). A good proxy is a measurable indicator strongly correlated with the intermittent factor (which is not directly observed.)[4]

If the link between the intervention and the outcome variable can be adequately represented by a variable, the effect of the intervention can be calculated by multiplying two parameters (the effect of the intervention on the intermittent factor and the effect of this factor on the final outcome). This indirect method, however, can only be applied if the following conditions are both met. First, the intervention affects the final outcome only through the intermittent factor, otherwise the impact would be underestimated. Second, the chosen intermittent factor is exogenous in the equation explaining the final outcome, i.e. all

[4] A good proxy should also remove any correlation between the unobserved variable and the other explanatory variables. E.g. the volatility of doctor-patient relationships is adequately measured by switching between institutions, and other factors explaining health outcomes do not add to this relationship.

other factors are controlled for (more precisely, the error term of the model and the explanatory variables must be uncorrelated). If the second condition is not met, the impact of the intermittent variable on the final outcome cannot be identified.

4.4 Explanatory variables in the regression

When building the model the inclusion of some explanatory variables should receive special attention.

- The development of the outcome variable in time should be captured by linear, quadratic or other polynomial trends. These trends can be different for different groups and can be captured by the interaction between dummies defining the groups and the trend. If the intervention variables show large cross-sectional variation then even yearly dummies can be used. (However, if there are interventions where the variance is purely time-based the impact of the intervention cannot be identified if yearly dummies are used and even the use of trends can be problematic if there are only a few years.) To control for the effect of seasonality within a year, it is recommended to use monthly or quarterly dummies in the regression.

- For example to control for the performance of different hospital units, hospital unit dummies can be used in the regression. Nevertheless, if these dummies are used then county dummies for example are not necessary because hospital unit dummies provide more detailed information.

- If dummies are used there should also be a baseline group to compare the behavior of other groups with. (Thus the number of dummies defined should be one less than the number of groups.)

- Together with (or instead of) a given explanatory variable, its function can also be included in the regression. For example, the effect of age is often modeled as a quadratic function. Another common transformation is the logarithm: e.g. instead of the change in the number of hospital beds, the change of the logarithm is often a more easily interpretable explanatory variable (which approximately equals to the percentage change for small changes).

4.5 Choosing the model and interpreting its parameters

Simple linear regression. Depending on the outcome variable of the regression different models should be chosen. For continuous variables the obvious choice is a linear regression model defined in the following way. Let y be the outcome variable, X_i (i=1,…,k) the explanatory variables (including the intervention variables and the constant as well) and β_i (i=1,…,k) the parameters to be estimated. In this case:

$$y = \sum_{i=1}^{k} \beta_i X_i + u \qquad (1)$$

where u is the unobserved error.

Interpreting the parameters. In a linear model, parameters are simple to interpret. Let X_1 be a dummy explanatory variable (e.g. indicating the intervention) and its parameter estimate β_1. Then the intervention changes the value of the outcome variable by β_1, controlling for other factors:

$$E(y \mid X_1 = 1) - E(y \mid X_1 = 0) = \beta_1. \qquad (2)$$

If the intervention variable is continuous, then β_1 shows the effect of a unit change in the outcome, controlling for other factors.

The easy interpretation of parameter estimates makes linear models easy to use. It must be noted however that linear regression is not always adequate even when the outcome variable is continuous. For example, hazard (duration) models work better if the outcome variable measures duration, or censored regression, if the outcome is censored. These models are not discussed in this brief note but are described in detail in the above cited econometric text books.

Logit and probit regression. If the outcome variable is a dummy (and this is the most common case in the indicator-based impact evaluation of healthcare interventions), then the probability of the occurrence of the outcome variable should be modeled. This is usually carried out with logit or probit functions. In this case the equation is:

$$P(y=1) = G\left(\sum_{i=1}^{k}\beta_i X_i\right), \quad (3)$$

where *P* denotes probability and *G* is a function of *x* with a range between 0 and 1. In the probit model, *G*=Φ, i.e. the standard normal distribution function, in the logit model *G* is the logistic function. In the following we desrcibe the logit model in detail as it is more commonly used in the medical literature. In this case the probality function takes the following form:

$$P(y=1) = \frac{e^{\sum_{i=1}^{k}\beta_i X_i}}{1+e^{\sum_{i=1}^{k}\beta_i X_i}}, \quad (4)$$

or equivalently:

$$\log\frac{P(y=1)}{1-P(y=1)} = \sum_{i=1}^{k}\beta_i X_i \quad (5)$$

The P/(1-P) ratio is known as the odds. The logit model therefore assumes a linear relationship between the log odds and the explanatory variables.

Interpreting the parameters. Probit or logit model parameter estimates are somewhat complicated to interpret as a percentage point change in the outcome induced by a dummy variable will also depend on the level of the other explanatory variables:

$$P(y=1 \mid X_1=1) - P(y=1 \mid X_1=0) = G\left(\beta_1 + \sum_{i=2}^{k}\beta_i X_i\right) - G\left(\sum_{i=2}^{k}\beta_i X_i\right). \quad (6)$$

In practice, the effect is evaluated at the mean values of explanatory variables (partial effect at the mean), or the average of the effects computed for each observation in the sample is reported (average partial effect).

In the logit model, equation (5) yields a simpler interpretation of parameter estimes. Assuming that one of the explanatory variables (e.g. the intervention variable) is a dummy and its estimated parameter is β, this would change the odds of the outcome variable approximately by $100\beta\%$ (exactly (e^{β}-1)–times). For example if the likelihood of occurrence (at the given

values of the other explanatory variables) is 20% (i.e. Odds=0.2/(1-0.2)=0.25) without the intervention and β=0.1, then after the intervention (all else remaining the same) Odds≅0.25*1.1=0.275, and this way the likelihood of occurrence has increased to approximately 0.275/(1+0.275)=0.216 or by 1.6 percentage points. However, if the initial likelihood of occurrence is 2% then the increase is only about 0.2 percentage point. It should be noted that if the likelihood of occurrence is small then the percentage change in the odds is close to the percentage change in *P*; therefore the intervention changes the likelihood of occurrence by about 100β% (but not by percentage points!). For example, in psychiatric clinical practice, the average share of patients readmitted within 30 days is only 2%, thus β=0.1 can be interpreted as the intervention increasing the average rate by 10% (to 2.2 from 2%).

Parameter interpretation is similar to the above if the explanatory variable is continuous. In the logit framework the estimated parameter shows that a (small) change of the variable changes the odds of the outcome variable by about 100β%. The exact change in percentage points can also be calculated at the average values of explanatory variables or the effect can be calculated for each observation and then taking their average.

If the intervention affects a continuous variable, the hypothetical value of the explanatory variable without the intervention should also be taken into consideration (for example how the number of hospital beds would have changed at the level of the counties). To determine the effect of the intervention predicted outcome variables should be calculated for each observation with and without the intervention and their average difference is the average partial effect of the intervention.

4.6 Estimating and testing parameters, model selection, goodness of fit and diagnostics

Estimating and testing the parameters. The commonly used statistical softwares automatically carry out the estimation of models (simple linear, logit, probit, duration) with the necessary techniques (least squares, maximum likelihood etc.); therefore the methods of estimation are not discussed here. In a simple linear model it is always best to use estimators with standard errors robust to heteroscedasticity (so called White's ***HETEROSCEDASTICITY***-consistent estimator), which is available in most statistical software. (Those who are interested can find the detailed description of econometric techniques in Woolridge (2002), or an intuitive explanation in Wooldridge (2009).)

The output of the estimation includes a point estimate for all parameters and the standard error. The ratio of these two is the t-statistic which gives the significance of each parameter: if the absolute value of the t-statistic is larger than 1.96 then the parameter is significant at 5% level, if it is larger than 2.58 the significance level is 1%.[5] Most softwares calculate the p-value: the lower the p-value, the more significant the parameter is. Generally a variable is considered significant if the p-value is smaller than 5%.

More than one parameter can be tested at the same time using a Wald-, an LM or a likelihood ratio (LR) test. These tests are automatic options in most types of softwares, however if not, the LR test can easily be calculated by hand. The general model and the model where all

[5] These values are valid for large samples, which is usually the case for available health care micro data.

parameters that are being tested are set at zero, are estimated, and the test statistic is twice the difference between the log-likelihood of the two models. (The log-likelihood value of the model is always included in the output.) This test statistic value should be compared to the 5% critical value of χ^2 –distribution where the degree of freedom equals to the number of tested parameters.

Model selection. The choice of the "best" model is more art than science. To start with, all potentially relevant variables can be included in the model and then gradually excluding variables or groups of variables that are not significant. If a group of variables (for example all county dummies) is significant as a whole, however many of the individual variables are not significant, it might be useful to merge them, for example groups of counties into regional dummies. Similarly the institutional dummies might have to be part-merged. Another modeling approach, however, suggests that the initial model should be narrower and only include those variables that definitely have an impact on the outcomes and are related to the intervention variable, and this model can be expanded to include – if necessary – other significant variables.

Overall, the final model should contain significant variables (or at least groups of variables that are jointly significant); although a few non-significant variables can also be left in the model, which are relevant from a (medical or economic) theoretical perspective and for the interpretation of the results. Even with this condition there are many potentially "good" models because a formal significance test is not possible for non-embedded models (when any model's set of variables is not a subset of another's). In these cases different model selection criteria can be used (e.g. Akaike or Bayesian information criterion). Finally, if more than one models seem "good" then it is worthwhile to check whether they yield similar results in terms of their main conclusions (the impact of interventions), i.e. if their conclusions are robust.

"Goodness of fit" analysis. In linear regression frameworks the most commonly used goodness-of-fit indicator is R^2 that shows to what extent the model explains the variance of the outcome variable (therefore its value ranges between 0 and 1). In logit and probit models a similar indicator, the pseudo R^2 is used,[6] or a more simple measure is the percentage of well-classified cases. To calculate the latter, the probability predicted by model (3) is calculated for each observation and it is assumed that the model predicts the 1 for the value of the outcome variable if the probability is larger than 0.5, otherwise it is 0. At the end, the percentage of correct outcome predictions is counted. This indicator can be misleading if the probability of one of the values (0 or 1) is very low, therefore sometimes the threshold is not set at 0.5 but at the outcome variable's average likelihood of occurrence.

Model diagnostics. In a simple linear model residuals (the error term) are important to examine. If the model specification is correct, then the conditional expectation of the residuals must be zero for each explanatory variable. If plotting an explanatory variable against the residuals reveals a *functional relation*, this suggests that the functional form of the estimated equation was misspecified (e.g. the true relationship is quadratic rather than linear) and the model should be changed accordingly. If the standard deviation of residuals is correlated with an explanatory variable (this is called *heteroscedasticity*), then we must use a standard error

[6] Definition: $1-L_{ur}/L_0$, where L_{ur} denotes the log-likelihood of the unconstrained model and L_0 denotes that of the constrained one containing only the constant.

estimator robust for heteroscedasticity (a built-in option in most statistical software). Residuals may also be correlated in time or space – this is called *autocorrelation*. This is a typical problem in panel or time series data, which can be solved by using standard error estimators robust for autocorrelation (e.g. the Newey-West estimator) or by eliminating the autocorrelation using quasi-differentiation. Often the *normal distribution* of residuals are tested as well, however, this is not very important in linear models as parameter estimates have favourable properties (are consistent and asymptotically normal) even without normally distributed errors if the sample is large enough.

Discrete dependent variable models. If the outcome variable is discrete, the above diagnostic problems (eg. heteroscedasticity) may show up in the misspecification of functional form of *G* (the probability funcion). The formal diagnostics of discrete dependent variable models are more complicated than of linear models, so in practical applications they are usually substituted by an easier solution. This is to estimate more than one specification (e.g. logit and probit) and accept results if average partial effects are similar across specifications.[7]

5. Summary of results and limitations

Statistical and "substantive" significance. The significance level must always be reported together with the calculations and interpretation of the impact of interventions and the economic ("substantive") significance must also be considered. In a large sample it is often the case that all of the variables are highly significant but only a few variables have substantive effect. The scale of the effects must always be clarified.

Endogeneity – exogeneity. As already mentioned above, it is important to emphasise in the discussion of the results that the estimated parameters can only be interpreted as a causal relationship if the explanatory variables are exogeneous, that is, all other relevant factors are controlled for. Exogeneity can be compromised by a number of related issues or put differently, endogeneity can be caused by various reasons.

One of the cases (already mentioned above) is when there are *unobservable variables* in the background that affect both the explanatory variable and the outcome variable. These variables – since they are unobserved – cannot be directly included in the regression. For example in the case of hospital re-admissions within 30 days, if one of the explanatory variables is the duration of hospital care then its coefficient (even if it is controlled for diagnosis-related groups and other variables) will not tell the exact *impact* (causal relationship) of the duration of hospital care on hospital re-admission. The more severe cases within the diagnosis-related groups are probably discharged later and are more likely to be readmitted within 30 days; nevertheless there is no exact measurement of "severity". For example, the method of instrumental variables can be used to address this problem: it aims to identify an exogenous "shock" that only has an impact on the outcome variable through the given explanatory variable (for more details see e.g. Wooldridge 2002 or 2009). In practice, the issue of unobservable variables should always be considered in the analysis.

Another important issue, particularly in healthcare-related applications, is *selection bias*. For

[7] It should be emphasised that the β parameters may be very different across specifications. What is important is that average partial effects should be similar.

example if the recovery from a certain illness is studied then data are available only for people who went to see a doctor and got a diagnosis while those who recovered without any medical intervention will never be included in the sample. Selection is non-random in this case and might be related to the symptoms: those with milder symptoms are less likely to see a doctor. Therefore the selected subsample is not representative of the total population and if the analysis is limited to this it might give biased results. The first step of the solution is considering the selection opportunities and as a next step, where possible, modeling the selection process. For more detail see the econometric text books (e.g. Wooldridge 2009). As with unobserved variables, selection bias must always be checked for.

6. References

Grun, R. (2006): Monitoring and evaluating projects, a step-by-step primer on monitoring, benchmarking and impact evaluation. HNP Discussion Paper, World Bank.

Khandker, S. R., Koolwal, G. B., Samad, H. A. (2010): Handbook on impact evaluation, quantitative methods and practices. World Bank.

Wooldridge, J. M. (2002): Econometrics of cross section and panel data. The MIT Press.

Wooldridge, J. M. (2009): Introductory econometrics, a modern approach. South Western Publishing.