WORK-IN-PROGRESS – PLEASE DO NOT CITE OR CIRCULATE

Can differences in Patient Reported Outcome Measures be attributed to providers?

Silviya Nikolova, Matt Sutton

Health Sciences – Economics, School of Community Based Medicine, University of Manchester, Jean McFarlane Building, Oxford Road, M13 9PL Manchester

Abstract

Patient Reported Outcome Measures are now collected routinely from patients undergoing four major surgical procedures in the NHS in England. Variations between providers are being interpreted as variations in their achievements. However, these variations also represent differences between providers in both observable and unobservable differences in case mix of patients and true random variation. Using the national PROMs data and two other datasets containing patient reported outcomes, we consider whether these variations can be attributed to providers with confidence. We find that provider variation is reduced substantially by allowing for observable case-mix, particularly baseline health, by increasing the number of follow-up observations on patients, by allowing for patient-level unobservable heterogeneity, and by estimating provider variation on a greater number of patients. These findings suggest caution is warranted in interpreting variations in short-term PROMs between providers.

Paper presented to CES-HESG, Aix-en-Provence 11-13 January 2012

Introduction

Since April 2009, Patient Reported Outcome Measures have been collected from all patients receiving four procedures in England before and after undergoing treatment. Once case-mix adjusted, differences between providers are being interpreted as indicators of variations in the quality of care. These provider variations are being used to benchmark providers and there are proposals that provider achievements of outcomes may be linked to payment.

The changes in outcomes that are observed reflect the effects of a wide range of factors, including observable and unobservable patient variables and the contributions of health care providers other than the hospital providing the procedure of interest. Adjustment of the outcomes using observable measures of case-mix may not be sufficient for residual differences between hospitals to be attributable to variations in care quality.

This problem of sorting on unobservables was analysed by Doyle et al (2010) using a unique natural experiment in which nearly 30,000 patients were randomly assigned to two different clinical teams in a large, urban Department of Veterans Affairs (VA) hospital. They found that costs differed between teams but the health outcomes (measured by 30-day readmissions, as well as 1-year and 5-year mortality) were the same. We consider more refined measured of health outcomes than these proxy markers.

Lilford et al (2004), among others, highlight the problems with interpreting observed variations in outcomes across providers noting that most studies fail to find close links between quality markers and patient outcomes.

This problem of attribution is similar to the problems faced by labour economists in the analysis of variations in wages between employees and between employers. Econometric models that control for unobservable employee-level heterogeneity have been developed to isolate the effects of employer heterogeneity on wages. We consider the applicability of these concepts and models to Patient Reported Outcome Measures. We use three data-sets containing PROMs that have different properties in terms of the health conditions, the interventions, the types of providers and the number of patient follow-up observations.

Data

We make use of three datasets with different properties: the national PROMS, a longitudinal register of patients diagnosed with rheumatological conditions and a panel survey of individuals living in private households.

National Patient Reported Outcome Measures data

The Patient Reported Outcome Measures (PROMs) are patient-level data collected from all providers of NHS-funded care. The data have been collected routinely since the 1st April 2009. Patients are surveyed before and after surgery, using paper-based self-completion questionnaires. These four large-volume procedures are hip replacement, knee replacement, hernia repair, and varicose veins. For each operative procedure the questionnaire collects information on EQ-5D, a widely used generic (disease non-specific) quality of life measure. The EQ-5D self-completion questionnaire asks patients to classify themselves as having one of three levels of health – no problems, some problems or extreme problems – in each of five dimensions of health – mobility, self-care, usual activities, pain/discomfort and anxiety/depression. This results in an EQ-5D health profile for a patient. The health utility score can be measured by the EQ5D score generated for each patient by assigning to their EQ5D profile its social value (Measuring and Valuing Health Study (Dolan, 1997)). The PROMs initiative also collects information on the visual analogue scale, the EQ-VAS, which records each patient's overall assessment of their health on a scale from 100 (best imaginable health) to 0 (worst imaginable health).

The PROMs also include information on clinical characteristics (frequency and duration of symptoms; previous surgery, co-morbidities, and general health) and socio-demographic characteristics (age, sex, and index of deprivation based on the respondent's postcode). Post-operative questionnaires also collect information on patients' views about the results of their surgery and post-surgical complications.

The national PROMs data for our analysis cover the period up to November 2010. We exclude observations for which the status is incomplete[1]. This reduces the initial sample of 171,505 patient-level observations to 102,609. Each of these patients has provided complete data for the two periods of observation: before and after surgery. We further eliminate 192

[1] The status of the record is complete if both questionnaires 1 and 2 have been completed.

observations with duplicate episode identifiers. The final data set contains 102,417 observations.

We next merge the National PROMs data set to the Hospital Episode Statistics (HES) for 2009/2010. There are 2005 observations included in the PROMs data set which have been collected as part of the Patient Outcomes in Surgery (POIS) survey. Because our goal is to track the attribution of different level of provider care to patient health outcomes and the PROMs data provides information only on the postcode of the provider, but does not identify the consultant, which is routinely collected in HES, we exclude all POIS observations. Our final analytical sample contains 100,412 person level observations. The merged data set has a hierarchical structure with patients assigned to 4400 consultants and consultants clustered within 385 hospital sites.

British Society for Rheumatology Biologics Register

The British Society for Rheumatology (BSR) established a register of patients newly treated with biological agents, the BSR Biologics Register (BSRBR), which became operational in January 2002. The goal is to register all patients in the United Kingdom with rheumatic diseases, newly starting treatment with these agents and to follow them to determine any short and long term hazards to health. The BSRBR is a hierarchical data set where patients are under the supervision of consultants and consultants are clustered within hospitals. There is no movement recorded in the dataset over time of patients between consultants and consultant between hospitals.

At baseline all patients with rheumatic disease commencing therapy are asked to provide information on the individual components of their disease activity score (DAS 28), details on their previous and current therapy, and comorbidity. Data are also collected on smoking habits and occupational history, consultant's name and their hospital association. For three years patients are surveyed every six months and complete the EQ5D questionnaire giving a maximum of seven EQ-5D data points per individual. The responses to the EQ-5D are used to assign a preference-based societal utility estimate to the respondent's EQ-5D profile as above. They also provide information on current disease activity and the development of any adverse events. There are 45,301 patients clustered within 518 consultants and 257 hospitals.

British Household Panel Survey

The British Household Panel Survey (BHPS) is a longitudinal survey of individuals living in private households in Great Britain. This includes rich information on occupational, socio-demographic and health variables. The BHPS was designed as an annual survey of each adult (16+) member of a nationally representative sample of more than 5000 households, with a total of approximately 10,000 individual interviews. The same individuals are re-interviewed in successive waves and, if they split off from their original households are also re-interviewed along with all adult members of their new households. In our analysis we use an unbalanced panel from all 18 waves of BHPS data.

At each wave the BHPS collects responses to the 12-item version of the General Health Questionnaire (GHQ-36). The GHQ is a self-administered screening test aimed at detecting psychiatric disorders that require clinical attention The GHQ-12 gathers information on concentration, sleep loss due to worry, perception of role, capability in decision making, whether constantly under strain, perception of problems in overcoming difficulties, enjoyment of day-to-day activities, ability to face problems, loss of confidence, self-worth, general happiness and whether suffering depression. Respondents rate each item on a four-point scale (ranging from 0 to 3, 0 being the best score). A Likert scale is used to form an overall score across the item specific responses. This provides a mental health measure ranging from 0 (least distressed) to 36 (most distressed). The predictive validity and content validity of the GQ are good in comparison to other well-known scaling tests of mental illness (Turner-Bowker et al., 2002).

We obtained Primary Care Trust (PCT) identifiers for the main BHPS data set and included English residents only. The linked data set has a hierarchical structure where individuals are clustered within PCTs. The final data file has 146,765 person-level observations clustered within 152 PCTs.

Methods

We used hierarchical linear models to account for the clustering of patients within higher-level units. In hierarchical data settings, all n observations are organized within M independent groups (clusters). Assume y_{ij} corresponds to the health outcome variable in the j-th cluster, j=1,.., M, for cluster j consisting of n_j first-level observations i, $i = 1, \dots, I_j$.

$$y_{ij} = \beta X_{ij} + Z_j u_j + \varepsilon_{ij} \qquad (1)$$

There are M unobservable higher-level effects u_j, X_{ij} observable characteristics and ε_{ij} is the random component.

The clustered-data representation of the mixed model given in (1) can be extended to two nested levels of clustering, creating a three-level model once the clustering of patients within consultants, and consultants within hospitals is considered. In the PROMs and BSRBR data sets, provider effects are both at the consultant and hospital Trust level. Assume y_{ijk} is the health outcome variable for $i = 1, \dots, I_{jk}$ first-level observations nested within $j = 1, \dots, J_k$ second-level groups, which are nested within $k = 1, \dots, K$ third-level groups.

$$y_{ijk} = X_{ijk}\beta + Z_k^3 u_k^3 + Z_{jk}^2 u_{jk}^2 + \epsilon_{ijk} \qquad (2)$$

Z_{jk}^3 is the matrix for the third-level unobservable effects u_k, and u_{jk} is for the second-level unobservable effects.

The presence of movements of individuals between higher-level units (PCTs) in the BHPS data allows us also to analyse higher-level attribution using a three-way error component model or Fixed-Effects Least-Squares Dummy Variables Regresssion (FELSDVREG). The procedure is based on the original work of Abowd, Kramarz & Margolis (1999). The model estimates the fixed effects for PCTs, patients and periods by setting all unidentifiable effects to 0. In the BHPS data, problems with FELSDVREG identification could arise if there is only one patient in a given PCT. In this case the individual effect will not be identified, which, however, is an unlikely scenario given that PCTs are large geographical units. It is possible also that there are no individuals moving in and out of a PCT, in which case the PCT effect is not identified. Formally,

$$y_{it} = \theta_i + \psi_{j(i,t)} + X_{it}\beta_x + \varepsilon_{it} \qquad (3)$$

where y_{it} is an observation for individual i, and time t, with $i = 1, \dots, N,\ t = 1, \dots, T$. θ_i is the fixed effect for individual i. $\psi_{j(i,t)}$ is the provider $j(i,t)$ fixed effect. X_{it} is a vector of observable, time-varying, exogenous characteristics of individual i. ε_{it} is the statistical residual with conditional mean zero. The three-way error component model has previously

been applied to isolate employee-level heterogeneity from the effects of employer heterogeneity on wages.

Multi-level models (`xtmixed` in Stata11) with random intercepts and no explanatory variables were used to partition the variability for each health outcome measure of interest by level (consultant, hospital Trust and PCT where applicable). The national PROMs data were analysed by procedure[2]. The estimates from the variance-covariance matrix were used to calculate the intra-class correlation for each level in the hierarchy, which indicates the amount of variance in the dependent variable that is attributable to each level in the model. Exploratory analysis regarding the distribution of the dependent variable reveals that the normality assumptions underlying the hierarchical level model are not met when the EQ-5D measure is the health measure of interest. However, the primary goal of this analysis is to demonstrate the usefulness of the different econometric techniques in analysing provider attribution to patient health outcomes under different data structures. We repeat our analysis using the EQ-VAS measure available in the National PROMs data. To match the BSRBR results against the PROMs findings we run regressions using one period of follow-up data in the BSRBR. We have further graphically explored how the extent of variation in provider effects changes as more waves of data are incorporated into the analysis.

To relax the assumption in the random effects formulation of the distributions of the higher-level effects, and their orthogonality to the other variables, when we have three levels we first run a model with consultant fixed effects. Then we save them and use them as dependent variables in a second stage regression which contains fixed effects for hospitals.

We next considered the inclusion of patient level characteristics as adjusters for case-mix. In all our analysis we control for age, sex, and comorbidities. When analysing mental health using the BHPS data, we have included income, educational attainment, number of children for three age groups, and a dummy variable for whether the person is living alone which have been identified as significant health determinants in Contoyannis et al (2004). The body-mass index (BMI) has been added to the analysis of health outcome for rheumatology patients in the BSRBR data set[3]. This analysis permits an assessment of how much variance in health outcomes can be explained by commonly used adjusters that are observable to the econometrician.

[2] The National PROMs collects information on hernia, hip replacement, knee replacement, and varicose veins .
[3] In the PROMs we include age, sex, and comorbidities to risk-adjust the health outcome of interest. In the BSRBR we include age, sex, comorbidities, and bmi. In the BHPS we control with income, educational attainment, number of children for three age groups, and a dummy variable for whether the person is living alone in addition to age, sex, and comorbidities.

We have further sought to identify the gain of including baseline health measure in explaining the variance. First, we treat baseline health as a measure of the baseline health endowment. This, however, could be endogenous if affected by individual and provider choice. Second, we implement the differencing procedure proposed by Todd and Wolpin (2003).

$$\Delta y_{i,j,t} = \Delta X_{i,j,t}\alpha + \gamma \Delta y_{i,j,t-1} + \nu_{i,j,t}$$

where $\Delta y_{i,j,t-1}$ is lagged health. According to them, in an optimising behavioural model, baseline achievement (here, baseline health) is expected to affect the choices made by individual and providers and, therefore, baseline achievement has persistent effect on outcomes in future time periods. For linear models with additive time-invariant unobserved effects for individual and provider, it is possible, through a simple differencing procedure, to obtain consistent estimates of the parameters if there exists a third (earlier) observation on achievement, along with data on the covariate set $X_{i,t-1}$. Todd and Wolpin specify two additional assumptions that must be met for the differencing procedure to produce consistent estimates: a) initial health endowment is fixed for life at conception and b) the effect of other inputs is not age-specific.

Finally, we inquire how controlling for person level time-varying characteristics changes the distribution of consultant and medical centre provider effects.

Results

Descriptive statistics

The individuals in our BHPS sample have an average GHQ score of 11.1 with standard deviation. At baseline individuals who join the BSRBR have an average utility score of 0.32 with a standard deviation of 0.34. Their health scores are stable over time and less variable during the follow-up periods. The mean utility score for the six periods of subsequent observation is 0.49.

All PROMs patients at baseline have an average EQ-5D score of 0.51 and standard deviation of 0.34. The average score for the second PROMS period is 0.78 and its standard deviation is smaller. Hip and knee patients have lower baseline and lower post-surgery scores than patients with varicose veins or groin condition. The variability in the outcomes of hip and knee patients is also larger. There is significant variability in the pre-surgery EQ-5D scores

for hip and knee patients at either level of the system. There is still variability in the post-surgery scores though the gap between the 25^{th} and 75^{th} percentile scores gets smaller. For the case of vein and groin surgeries there is some variation at either level of the system, though the scores are very much the same for consultant and hospital facility level. This is true for both periods of observation. Similar observations can be made regarding the EQ-VAS health measure. The summary statistics in Table 1 show that there is variability in the health metrics at the consultant and hospital levels and over time within these levels and that health outcome means increased following surgical intervention (as in PROMs and BSRBR data).

Proportion of variation explained by health care provider

Of the total variation in the patient health outcome variables (GHQ-36 and EQ-5D) at most 10.6 % was explained by health care providers. Table 2 reports the percentage of total variation at the provider level and the percentage of provider level variation at the two different levels. In the BSRBR data set, the hospital level explained the largest share of the provider-level variation. In the PROMs data we find the consultant level explains 73.2 % of the total provider variation. A separate examination of each PROMs procedure shows that the percentage of total system variation explained is at most 3.5% (hip surgery). For all procedures, the hospital level explained the largest share of provider variation. For the EQ-VAS health outcome variable, consultant and hospital variation jointly explains 4.9% of the total variation with consultant contributing 59.3% to it. Analyzing EQ-VAS measure by procedure we find the consultant and medical facility levels explain jointly at most 2.5% of total variation (knee surgery). Consultant attribution to system level variation is larger for vein and groin procedures.

Adjusting for person-level characteristics

Adjusting for person-level characteristics reduces the maximum percentage of total variance explained to 6.5% (Table 2). We obtain conflicting results as to which level of the delivery system explains the largest share of the variation.

Including baseline health

Including baseline health reduces the maximum percentage of total variance explained by the system to 2.4 %. In the BSRBR we find that the hospital level explains 65% of provider variation. In the PROMs data variation across hospitals explains a larger share (69%) of EQ-5D variation. For hip surgery, knee surgery, and varicose veins we find that hospital explains respectively 71.2%, 91.5%, and 100% of total provider variation. In the case of groin surgery our results show that variability between consultants explains 0.4% of the total variability in the system. Consultants and hospitals jointly explain at most 1.7% of the variation in the EQ-VAS measure with hospitals explaining a larger share of it in the case of hip and knee surgery (53.8% and 71.2%).

The results in Table 2 also include the findings from the differencing procedure as proposed by Todd and Wolpin (2003). We find that, once all time-invariant characteristics are differenced out, there is no variability between consultants, hospitals, and in trends between providers and hospitals in the original data.

Allowing for individual-level unobservable heterogeneity

Results from the three-way error component model applied to the BHPS data show that PCTs explain only 0.2% of total variation (Table 3). At most 9.3% of the total variation is explained by the time-varying person-level characteristics controlled for in the estimation. However, most of the variation (43% for standard case-mix adjustment model or 39.2% for a model including baseline health as well) is from stable individual characteristics (chronic diseases, habits, or risk factors, for instance, elevated high-blood pressure). Approximately half of total variation is explained by random variation (ε_{it}).

Fixed effects estimation of the higher level effects

Our results in Table 4 show that hospital fixed effects jointly explain 58% of the variation among consultants in the BSRBR. Consultant effects and mix-adjusting variables explain 9.4% of total variation in utility among patients. In the National PROMs, hospital fixed effects explain at most 26.9% of the variation among consultants (knee surgery). At the same time consultants explain only 5.2% of the utility score variation among patients who underwent knee surgery. Our results are very similar for hip surgery. For groin and varicose vein procedure hospitals explain a smaller share of total variation. We find that, in the case of

groin surgery, consultants explain 17.9% of the variation in the EQ-5D utility score. Our results for the EQ-VAS measure point to a larger share of consultant and hospital variation explained for groin and varicose vein procedures.

Effects of adjusting for observable differences in case-mix

The variation occurring at each level of medical care delivery was not uniformly affected by standard case-mix adjustment variables (Table 5). In the BHPS data their inclusion reduced variability across PCTs by 36.3%. The variation across consultants in the BSRBR data set was completely eliminated, whilst facility variation remained the same. PROMs results point to 61.4% decline in variation across consultants and to 29.1% decline in variation across hospitals. There were differences in the effect of case-mix adjustment depending on procedure. It reduced the provider variation for varicose vein and groin surgery (36.5% and 44.7% respectively). Consultant variation declined by 40.5% for vein and 9.6% for groin surgery. For hip surgery and knee surgery, system variation went down by 15.4% and 12.8% respectively. In the case of EQ-VAS measure case-mix adjustment reduced provider variation the most for vein and groin surgery (54% and 52.2% respectively). With the exception of EQ-VAS measure for varicose veins, we find that standard case-mix adjustment explains relatively little of the overall variation. Overall variation in the system declined by 11.7% at most (groin).

Effect of including baseline health

Including baseline health explains 88.8% of the variation across PCTs in the BHPS. Using all six follow-up periods of BSRBR data we find that adjusting for baseline health and standard case-mix variables eliminates 13.9% of the variability across consultants and almost all of the variability across hospitals. In the national PROMs dataset, baseline health along with standard case-mix variables could explain all of the variation across consultants in the case of varicose veins. It explains all of the variation between hospitals in the case of groin surgery. Total provider variation declined by as much as 28.9% at most. In the case of EQ-VAS measure total variation declined by as much as 34.5%.

Effects of case-mix adjusting on the distribution of consultant and medical provider effects

Results presented in Table 6 show provider effect means, standard deviations, 25th and 75th percentile values for an empty model, for a model with standard risk-adjusting variables and for a model with risk-adjusting variables and baseline health. We find that mean consultant effects, on average, increased, standard deviation decreased, and the distribution became more concentrated. Similar findings are observed at the hospital level.

The effect of limited follow-up of patients

BSRBR results for one period of follow-up show that controlling for standard case-mix eliminates all variability across hospitals and decreases the variance across consultants by 33.1%. The share of total variance explained is reduced by 39.5%. Baseline health in addition to case-mix adjusting variables reduced total provider variability by 24.4%.

The effect of incorporating more waves of data

Figure 1 shows the relationship between the estimated consultant effects from an empty multi-level model and the sample of patients on which those consultant effects are estimated[4]. Variability among estimated consultant effects does not vary across follow-ups. It implies that the distribution of patients remains approximately the same over time. In Figure 2 we observe a similar pattern for hospitals.

The effect of more patient observations

There are two effects associated with larger number of patients per provider. First, more patient observations per provider should, in theory, help us draw improved inferences about provider attribution to health outcomes. In general, having more observations decreases the variance of the estimated parameters. In the case of random effects these parameters are the respective shares of total variance explained by the different levels of the health care delivery system. Adding more patient observations does not affect the variance computed using random effects. It rather decreases the variance of these estimated variances. In Figure 3 we plot estimated provider standard errors as a function of the number of hip patients treated in the PROMs data. The plot shows a decrease in the standard errors of estimates as the number

[4] There are more patients joining the Registry over time.

of patients per facility grows large. A similar pattern is observed for all PROMs procedures as well as in all BSRBR waves. Second, large hospitals appear to deliver approximately the same quality of care (Figure 4). These hospitals, because of their large patient size, are less exposed to risk associated with sicker patients or lower quality consultant work.

Discussion

Even when case-mix adjusted for observables, variations across providers in Patient Reported Outcome Measures that are collected shortly before and shortly after treatment may be poor signals of variations in provider quality. We find that hospital variation, as a share of total provider variation, for one-off surgical procedures is smaller compared to medical treatment of rheumatic diseases over three years, and mental health over 18 years. At the same time, a small amount of health outcome variation is explained by providers. The three-way error component model results affirm the finding that small share of total variation is explained by time-invariant provider effects.

We find that provider variation is reduced by allowing for observable case-mix, particularly baseline health. The distribution of provider effects becomes more concentrated with mean consultant effects, on average, increasing, and the standard deviation decreasing. The three-way error component model results clearly demonstrate that person-level heterogeneity is most important in explaining provider-level variation in the BHPS. Results from the differencing procedure which relies on at least three data points show that eliminating time-invariant characteristics eliminates all variation between consultants, hospitals and in trends between consultants/hospitals.

Fixed effects estimation produce larger estimates of consultant variation than random effects estimation. The fixed effects model relaxes the assumptions of (i) orthogonality between the included variables and the higher-level effects and (ii) a normal distribution at the higher-level.

We observe that more patient observations per provider (i) improve the precision of statistical inference and (ii) big hospitals are approximately the same in terms of quality of medical care.

In future work, we will seek to adopt more sophisticated methodological approaches to account for the non-normality in the distribution of the EQ-5D and EQ-VAS variable and endogeneity of baseline health.

References:

Abowd, J., Kramarz, F., Margolis, D., 1999, High-Wage Workers and High-Wage Firms, Econometrica, 67, 2, 251-333.

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

Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. The Lancet Volume 363, Issue 9415, 3 April 2004, Pages 1147-1154.

Dranove, D., Daniel, K., McClellan, M., Satterthwaite, M., 2003. Is more information better? The effects of "Report Cards" on health care providers. Journal of Political Economy 111 (3), 555-588.

Dolan P. Modeling valuations for EuroQol health states. Med Care 1997; 35(11):1095-1108.

Doyle JJ, Ewer SM, Wagner TH. Returns to physician human capital: Evidence from patients randomized to physician teams, Journal of Health Economics 29 (2010) 866–882

Simmer, T.L., Nerenz, D.R., Rutt, W.M., Newcomb, C.S., Benfer, D.W., 1991. A randomized control trial of an attending staff service in general internal medicine. Medical Care 29, JS31-JS40.

Todd, P., Wolpin, I., 2003, On the Specification and Estimating of the Production Function for Cognitive Achievement, Economic Journal, Feb, F3-F33

Turner-Bowker, D.M., Bartley, P.J., Ware, J.E., Jr. *SF-36® Health Survey & "SF" Bibliography: Third Edition (1988-2000)*. Lincoln, RI: QualityMetric Incorporated, 2002.

Table 1: Summary Statistics for GHQ-36 and EQ-5D by Level of Care

	Patient			Consultant means		Hospital means	
	N	Mean	SD	p25	p75	p25	p75
BHPS	137062	11.1	5.33	n.a.	n.a.	7.57	13.63
BSRBR, baseline	3849	0.32	0.34	0.14	0.52	0.09	0.54
BSRBR, follow-up	32715	0.49	0.3	0.34	0.65	0.33	0.66
PROMs EQ5D, baseline	96100	0.51	0.34	0.41	0.71	0.31	0.76
Hip	29823	0.35	0.32	0.14	0.52	0.08	0.62
Knee	33606	0.4	0.31	0.21	0.58	0.13	0.67
Vein	9340	0.77	0.21	0.71	0.87	0.72	0.87
Groin	23331	0.79	0.2	0.72	0.89	0.74	0.93
PROMs EQ5D, follow-up	95116	0.78	0.25	0.71	0.91	0.69	0.97
Hip	29325	0.76	0.26	0.66	0.89	0.65	0.96
Knee	33084	0.7	0.27	0.62	0.83	0.62	0.87
Vein	9338	0.86	0.2	0.79	0.97	0.79	0.98
Groin	23369	0.87	0.19	0.79	0.97	0.79	0.99
PROMs EQ-VAS, baseline	89762	71.8	19.5	64.6	82.6	61.7	85.5
Hip	28102	66.2	21	56.1	76.7	53.2	80.1
Knee	31324	68.7	19.2	59.9	78.9	56.3	81.9
Vein	8754	80.1	15.7	73.7	89.3	74.5	90.3
Groin	21582	80.2	14.8	73.3	87.8	73.5	89.3
PROMs EQ-VAS, follow-up	95171	75.5	17.9	68.1	85.4	67.3	88.7
Hip	29201	75.2	18.3	67.2	83.9	65.7	88
Knee	33022	71.9	18.7	64.2	81.7	61.5	85.2
Vein	9405	79.6	16.2	72.8	89.2	72.8	90.1
Groin	23543	79.1	16	71.7	87.7	73.2	89.8

Table 2: Percentage of Variation Explained by Different Provider Levels

	Percentage of total variance explained by the provider level	Percentage of provider level variance explained by:	
		Consultant	Hospital
empty model			
BHPS	1.0	n.a.	100
BSRBR	6.1	3.9	96.1
PROMs, EQ-5D			
PROMs, hip	3.5	34.5	65.5
PROMs, knee	3.4	19.3	80.7
PROMS, vein	2.2	46.7	53.3
PROMs, groin	1.1	37.2	62.8
PROMs, EQ-VAS			
PROMs, hip	2.3	43.9	56.1
PROMs, knee	2.5	21.4	78.6
PROMS, vein	1.7	63.2	36.8
PROMs, groin	1.5	62.5	37.5
Adjusted for case-mix			
BHPS	0.7	n.a.	0.7
BSRBR	6.5	0.2	99.8
PROMs, EQ-5D			
PROMs, hip	3.1	33.6	66.4
PROMs, knee	3.1	21.1	78.9
PROMS, vein	1.5	43.7	56.3
PROMs, groin	0.6	60.9	39.1
EQ-VAS			
PROMs, hip	2.0	41	59
PROMs, knee	2.0	24.7	75.3
PROMS, vein	0.9	0	100
PROMs, groin	0.1	31.4	68.6
Adjusted for case-mix & baseline heath			
BHPS	0.1	n.a.	100
BSRBR	1.0	35	65
PROMs, EQ-5D			
PROMs, hip	2.4	28.8	71.2
PROMs, knee	2.4	8.5	91.5
PROMS, vein	0.8	0	100
PROMs, groin	0.4	0.4	0
PROMs, EQ-VAS			
PROMs, hip	1.7	46.2	53.8
PROMs, knee	1.2	28.8	71.2
PROMS, vein	0.2	100	0
PROMs, groin	n.a	n.a.	n.a
Differencing procedure			
BHPS	0		
BSRBR	0		

Table 3: Fixed Effect Least Squares Dummy Variable Regression using the BHPS

	Percentage of Total Variance Explained by			
	Time-Varying Characteristics	**Person FE**	**PCT FE**	**Estimated Random Error**
Adjusted for case-mix	5.1	43	0.2	51.8
Adjusted for case-mix and baseline. health	9.3	39.2	0.2	51.3

Table 4: 2-Step Fixed Effects – Percent of Explainable variation

	PROMs, EQ-5D		PROMs, EQ-VAS		BSRBR, EQ-5D	
	% utility variation expl. by consultants	**% consultant variation expl. by hospitals**	**% utility variation expl. by consultants**	**% consultant variation expl. by hospitals**	**% utility variation expl. by consultants**	**% consultant variation expl. by hospitals**
all proc	13.5	21	10.1	20.1	9.4	58
hip	6.9	26.9	7.5	17.8		
knee	5.2	25.7	7.5	17.8		
vein	17.9	12.9	13.3	38.7		
groin	8.5	7.8	12.8	20		

Table 5: Percentage Change in Variation Because of Case-Mix

	Overall	**System**	**Consultant**	**Facility**
Casemix				
BHPS	9.6	36.3	n.a.	36.3
BSRBR	5.7	0	100	0
BSRBR, 1fup	5	20.8	12.6	100
PROMs, EQ-5D				
PROMs, hip	4.5	15.4	17.6	14.3
PROMs, knee	3.1	12.8	4.6	14.8
PROMS, vein	10.1	36.5	40.5	33
PROMs, groin	7.4	44.7	9.6	65.6
PROMs, EQ-VAS				
PROMs, hip	6.1	21.7	26.9	17.6
PROMs, knee	5.6	24.6	13.2	27.7
PROMS, vein	185	54	27.1	100
PROMs, groin	11.7	52.2	76	12.4
case mix and baseline health				
BHPS	28.3	88.8	n.a.	88.8
BSRBR	44.9	90.5	13.9	93.6
BSRBR, 1fup	24.4	39.5	33.1	100
PROMs, EQ-5D				
PROMs, hip	12.7	39.1	49.2	33.8
PROMs, knee	13.7	38.8	73.1	30.7
PROMS, vein	28.9	71	100	45.7
PROMs, groin	21.1	69	16.7	100
PROMs, EQ-VAS				
PROMs, hip	16.9	41	37.9	43.3
PROMs, knee	20.6	62.1	49.1	65.7
PROMS, vein	34.5	90.3	84.6	100
PROMs, groin	n.a	n.a	n.a.	n.a.

Table 6: Distributions of Provider Effects Across Levels of the Delivery System

	Consultant				Facility			
	Mean	**SD**	**p25**	**p75**	**Mean**	**SD**	**p25**	**p75**
empty model								
BHPS					-0.0255	0.4309	-0.3430	0.2662
BSRBR	0.0001	0.0027	-0.0016	0.0018	0.0025	0.0564	-0.0367	0.0397
BSRBR, 1fup	0.0024	0.0310	-0.0143	0.0198	0.0003	0.0031	-0.0015	0.0019
PROMs, EQ-5D								
PROMs, hip	0.0019	0.0130	-0.0060	0.0104	0.0024	0.0221	-0.0105	0.0154
PROMs, knee	0.0004	0.0075	-0.0040	0.0050	0.0000	0.0270	-0.0164	0.0171
PROMS, vein	-0.0003	0.0087	-0.0051	0.0048	-0.0009	0.0099	-0.0071	0.0054
PROMs, groin	0.0002	0.0035	-0.0017	0.0024	0.0004	0.0055	-0.0024	0.0034
PROMs,EQ-VAS								
PROMs, hip	0.1022	0.8728	-0.3995	0.6655	0.0807	1.0107	-0.4921	0.7098
PROMs, knee	0.0149	0.4499	-0.2482	0.2700	-0.0032	1.4257	-0.8263	0.8489
PROMS, vein	-0.0237	0.7913	-0.3879	0.4325	-0.0386	0.4541	-0.2853	0.2551
PROMs, groin	0.0704	0.6704	-0.3403	0.4363	0.0136	0.0136	-0.1966	0.2301
case-mix and baseline heath								
BHPS					-0.0030	0.1354	-0.1033	0.0977
BSRBR	0.0006	0.0059	-0.0027	0.0040	0.0011	0.0101	-0.0047	0.0064
BSRBR, 1fup	0.0030	0.0248	-0.0110	0.0123	0.0000	0.0000	0.0000	0.0000
PROMs, EQ-5D								
PROMs, hip	0.0008	0.0077	-0.0039	0.0054	0.0012	0.0168	-0.0075	0.0105
PROMs, knee	0.0001	0.0024	-0.0013	0.0016	-0.0004	0.0220	-0.0132	0.0132
PROMS, vein	0.0000	0.0000	0.0000	0.0000	-0.0007	0.0077	-0.0041	0.0038
PROMs, groin	0.0000	0.0037	-0.0020	0.0023	0.0000	0.0000	0.0000	0.0000
PROMs, EQ-VAS								
PROMs, hip	0.0423	0.6258	-0.2935	0.4412	0.0326	0.6577	-0.3173	0.4231
PROMs, knee	0.0056	0.2894	-0.1668	0.1636	-0.0084	0.6247	-0.3546	0.3291
PROMS, vein	-0.0022	0.1931	-0.1035	0.1102	0.0000	0.0000	0.0000	0.0000
PROMs, groin	0.0850	0.6663	-0.3142	0.4467	0.0371	0.4854	-0.2422	0.2968

Figure 1: Consultant effects variability at consecutive follow-ups

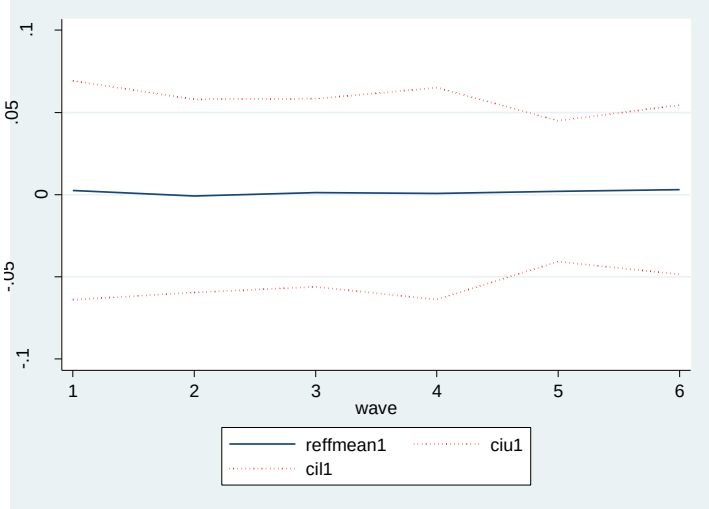

Figure 2: Hospital effects variability at consecutive follow-ups

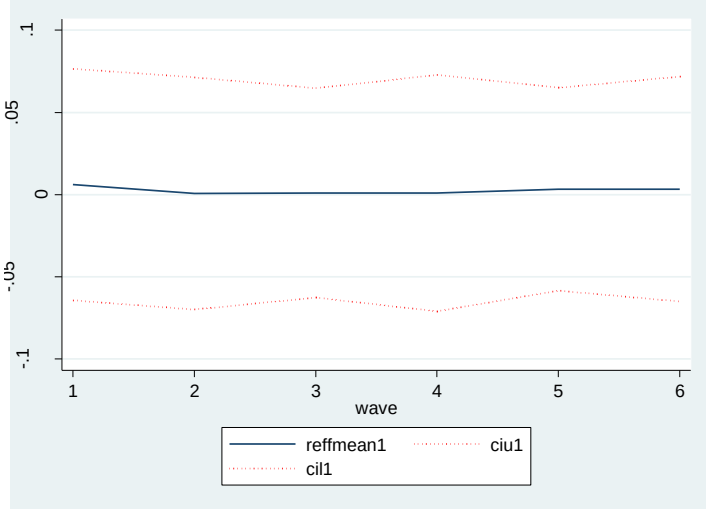

Figure 3: Standard errors of hospital effects and number of patients per facility

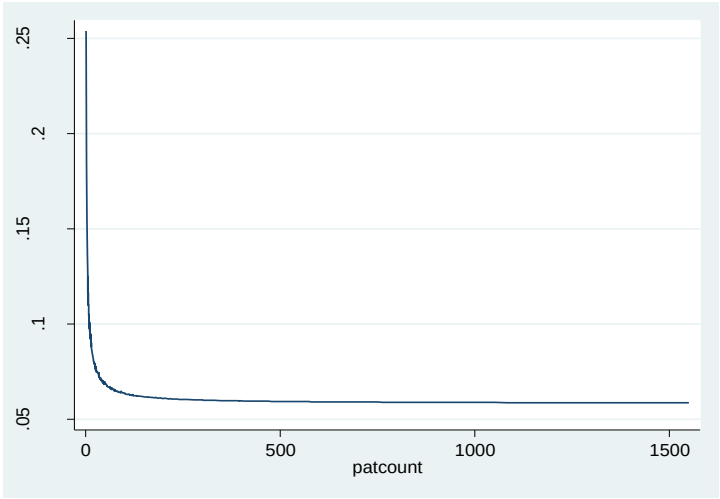

Figure 4: Hospital effects and number of patients per facility