

RESEARCH

Effect of telephone health coaching (Birmingham OwnHealth) on hospital use and associated costs: cohort study with matched controls



OPEN ACCESS

Adam Steventon *senior research analyst*¹, Sarah Tunkel *director*², Ian Blunt *senior research analyst*¹, Martin Bardsley *head of research*¹

¹The Nuffield Trust, London W1G 7LP, UK; ²Ernst & Young, London SE1 2AF, UK

Abstract

Objectives To test the effect of a telephone health coaching service (Birmingham OwnHealth) on hospital use and associated costs.

Design Analysis of person level administrative data. Difference-in-difference analysis was done relative to matched controls.

Setting Community based intervention operating in a large English city with industry.

Participants 2698 patients recruited from local general practices before 2009 with heart failure, coronary heart disease, diabetes, or chronic obstructive pulmonary disease; and a history of inpatient or outpatient hospital use. These individuals were matched on a 1:1 basis to control patients from similar areas of England with respect to demographics, diagnoses of health conditions, previous hospital use, and a predictive risk score.

Intervention Telephone health coaching involved a personalised care plan and a series of outbound calls usually scheduled monthly. Median length of time enrolled on the service was 25.5 months. Control participants received usual healthcare in their areas, which did not include telephone health coaching.

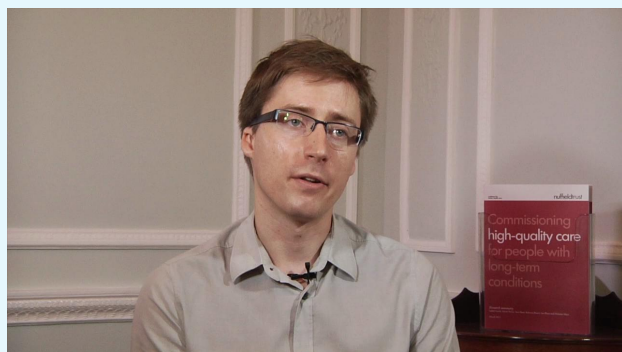
Main outcome measures Number of emergency hospital admissions per head over 12 months after enrolment. Secondary metrics calculated over 12 months were: hospital bed days, elective hospital admissions, outpatient attendances, and secondary care costs.

Results In relation to diagnoses of health conditions and other baseline variables, matched controls and intervention patients were similar before the date of enrolment. After this point, emergency admissions increased more quickly among intervention participants than matched controls (difference 0.05 admissions per head, 95% confidence interval 0.00 to 0.09, $P=0.046$). Outpatient attendances also increased more quickly in the intervention group (difference 0.37 attendances per head, 0.16 to 0.58, $P<0.001$), as did secondary care costs (difference £175 per head, £22 to £328, $P=0.025$). Checks showed that we were unlikely to have missed reductions in emergency admissions because of unobserved differences between intervention and matched control groups.

Conclusions The Birmingham OwnHealth telephone health coaching intervention did not lead to the expected reductions in hospital admissions or secondary care costs over 12 months, and could have led to increases.

Correspondence to: A Steventon adam.steventon@nuffieldtrust.org.uk

Video on bmj.com (see also <http://bmj.com/video>)



Video abstract

Introduction

Facing rising costs, healthcare systems around the world are exploring innovative ways to improve efficiency. Particular attention has been placed on the use of technology to help manage long term health conditions,¹ including one-to-one telephone health coaching. This involves a regular series of phone calls between patient and health professional. The calls aim to provide support and encouragement to the patient and promote healthy behaviours such as treatment control, healthy diet, physical activity and mobility, rehabilitation, and good mental health.² The hope is that the patient will maintain their own health more independently, and that the professional and patient will be in a better position to identify problems before they become critical. In turn, admissions to hospital may be prevented.³ Avoidable hospital admissions are both undesirable for the patient and expensive for the payer.

In a systematic review of telephone health coaching for people with long term conditions, only nine of 34 studies investigated effects on health service use.² Four studies showed effects in this area, but findings were hard to generalise because the studies looked at a range of different health conditions, had relatively small samples (the average sample was fewer than 360 people), and included interventions that were heterogeneous. Five of the nine interventions included telemonitoring of vital signs such as blood pressure alongside telephone coaching. Since the review, other larger studies have been conducted.

Wennberg and colleagues conducted a large randomised controlled trial of 174 120 patients with employer based health insurance.⁴ The intervention was given to patients who were at high predictive risk (for example, of future hospital costs), with a lower risk threshold used in one treatment group than in the other. The researchers concluded that telephone care management reduced hospital admissions overall and among patients with selected long term conditions (heart failure, coronary artery disease, chronic obstructive pulmonary disease, diabetes, and asthma), although there was no statistically significant effect on the subset of admissions that came through the emergency room. Among the subgroup of patients with long term conditions, overall medical and pharmacy costs were \$51 (£33.8; €38.8) per month lower in the more aggressively targeted group than the other. Their intervention did not include telemonitoring, but it did include an element of shared decision making for “preference sensitive conditions” (for example, with regards to treatment options for arthritis of the hip). The shared decision making could be largely responsible for the intervention’s effect on admissions that did not come through the emergency room.⁵

Further, Lin and colleagues studied telephone health coaching for 874 US Medicaid members of working age (range 18-64 years, mean 45.3) with at least one of 10 qualifying long term conditions and at least two acute hospital admissions or emergency department visits within a 12 month period.⁶ Their intervention relied solely on phone calls and mailed educational materials. Health coaches provided information on conditions and treatment options, empowered patients to self manage and self monitor their conditions, and encouraged patients to communicate their preferences to providers. Compared with a matched control group, the authors found no effects on hospital use and expenditures over one year. However, over two years, the number of emergency department visits reduced by a smaller amount for intervention patients than for matched controls, leading to a relative 20% increase in emergency department visits for intervention patients. Findings over one year were

later reproduced in a randomised controlled trial among Medicaid patients.⁷

There is continued interest in telephone health coaching with several providers, but the evidence base is unclear. As only a limited number of large studies have examined the effects on hospital use and have produced contradictory findings,⁴⁻⁶ more information is needed to understand the elements that make up a successful service. We were commissioned by the Department of Health in 2010 to evaluate the effect of England’s largest example of telephone health coaching (Birmingham OwnHealth) on hospital use and associated costs. Previous evaluations of Birmingham OwnHealth had shown that patients had high levels of satisfaction and believed that the service reduced their need to go to hospital.⁸ In addition, a study showed reductions in levels of glycated haemoglobin (HbA_{1c}), blood pressure, and body mass index among a subset of patients with poorly controlled diabetes.⁹ We present estimates of the effect of the Birmingham OwnHealth service on hospital admissions and associated costs. The service was decommissioned in 2012 after a consultation exercise.¹⁰

Methods

In summary, a retrospective design was used to assess the effects of an existing service. Matched controls aimed to reflect the changes in hospital use that can occur over time even without an intervention.^{11 12}

Intervention, including patient recruitment

Birmingham OwnHealth was established in 2006 in a large city with industry. This area had health inequalities and some parts with high deprivation. Birmingham OwnHealth aimed to improve self care strategies, improve clinical indicators, and reduce health service use.³

The service targeted people with heart failure, coronary heart disease, diabetes, or chronic obstructive pulmonary disease. Inclusion criteria were all of the following:

- A recorded diagnosis of one of the targeted conditions
- A minimum level of disease severity (for example, HbA_{1c}>7.4 in the past 15 months)³
- Age 18 or older
- Ability to communicate on the telephone
- A recorded address and practice registration.

Potentially eligible patients were identified through analysis of data extracts sourced from the participating general practices. Summary files were then reviewed and ratified by general practitioners. General practitioners applied additional clinical judgment in determining which patients to refer into the Birmingham OwnHealth service, based on a set of consideration factors. These factors included comorbidities under active treatment, personality and mental health problems, and life circumstances such as pregnancy. General practice staff sent an introductory letter to the selected patients, which was followed by a phone call by a representative from Birmingham OwnHealth.

Once enrolled, patients (who were known as “members”) were assigned a care manager, who were specially trained nurses employed by NHS Direct. General practice data were also transferred onto the operational systems used by Birmingham OwnHealth. During the programme, care managers followed five fundamental steps: assessment, recommendation, follow-up, ongoing management, and review.

Care managers made regular telephone calls to patients. These calls were usually made monthly at a predetermined date and time to suit the user, although a minority of patients received calls more frequently (two to four times per month) owing to disease severity, social isolation, or severe weather. As the number of members grew, some members were stepped down (or “graduated”) to quarterly calls. During the telephone calls, care managers asked patients about current health status and symptoms, and recorded this information along with other information such as recent test results and changes to treatment. Care managers then gave personalised guidance and support, aiming to build continuing relationships with patients and provide motivation, skills, and knowledge to encourage patients to better manage their health conditions. The calls focussed on eight priorities in care management—namely, to ensure that patients were able to do all of the following:

- Know how and when to get help for health and social care problems
- Learn about their condition, and agree and set treatment goals within a personalised care plan
- Take medicines correctly
- Get recommended tests and services
- Act to keep the condition in good control
- Learn how to make changes to lifestyle and circumstances to reduce risks
- Build on strengths and overcome obstacles, while strengthening personal social networks
- Follow up with specialists and appointments.

Calls were structured into modules focussing on each of these priorities, and used screen prompted algorithms to structure the conversation. The software prompted care managers to follow guidelines in each priority area (for example, provide basic information about treatment). Care managers, however, were not constrained to follow protocols. Additional educational materials could be sent to patients.

Care managers aimed to coordinate input across services, for example, when developing personalised care plans, and could refer patients onto existing services such as mental health services and social care. General practitioners were offered monthly phone calls and quarterly meetings with their assigned care manager to discuss patients. The service was set up to provide proactive calls to patients rather than act as a phone-in service, and inbound calls were less common, comprising about 5% of calls.

Study populations

Study participants were enrolled in Birmingham OwnHealth between the time the intervention commenced in April 2006 and December 2008. This cut-off period was chosen to ensure sufficient time to follow-up for at least 12 months when this study was commissioned. The service provider identified all such patients recorded in their operational datasets, from which we excluded those without a record of inpatient or outpatient hospital use in the three years before enrolment. Because the matching variables came from hospital data, we could not accurately characterise patients without previous hospital activity. However, we would expect these excluded patients to have low future levels of hospital use,¹¹ and therefore represent little potential to reduce hospital costs over 12 months.

The large size of the service reduced the scope to find matched controls locally, and there may have been spillover effects. Therefore, we selected several comparable areas within England

to provide a pool of potential matched controls. Four of these areas (Bradford and Airedale, Sandwell, Stoke, and Wolverhampton) were drawn from a national area classification,¹³ which were similar to the intervention area in terms of demography; occupational mix; and rates of education, occupation, limiting long term illness, and unpaid care. We also included Walsall, which was commonly used as a comparator by the Primary Care Trust that commissioned Birmingham OwnHealth. We checked that a similar health coaching service via telephone had not operated in the selected areas, using internet searches and discussions with colleagues. As a result, we excluded parts of Walsall from the pool.

Study endpoints and sample size calculation

Our primary endpoint was the number of urgent and unplanned (“emergency”) hospital admissions per head over 12 months. Our hypothesis was that the service could alter rates of emergency admissions in either direction. Increases in emergency admissions have been suggested by studies of other types of interventions involving patient outreach in England.¹⁴

We performed a sample size calculation at the outset of the study to check that we were likely to have data for a sufficient number of patients to produce meaningful conclusions. We thought it important to detect relative changes of 15% should they occur, based on the level of effect judged as meaningful in similar studies,¹⁵ at 90% power and with two sided $P=0.05$. Annual admission rates for the usual care group were assumed to be 0.25 per person with a standard deviation of 0.4.¹⁶ Calculations were performed in SAS 9.3, and assumed a correlation of 0.15 between the number of admissions for intervention and matched control patients. Based on these assumptions, 2035 intervention patients were needed.

Secondary endpoints calculated over 12 months included the number of planned (“elective”) hospital admissions, number of hospital bed days, number of attendances for hospital based ambulatory care (“outpatient attendances”), and costs of secondary care.

Data sources and data linkage

The service providers had access to identifiable data for participants, including a national patient identifier (the “NHS number”), sex, date of birth, and postcode. These data were transferred to the NHS Information Centre for health and social care who used them to link participants to national administrative data for secondary care activity (the hospital episode statistics (HES)).¹⁶ The data linkage required an exact or partial match on several of the variables at once. After the data had been linked, the HES identity was transferred to the evaluation team together with the date of patient enrolment into Birmingham OwnHealth, year of birth, sex, and small geographical area code. As a result, we only had access to “pseudonymised” data in which all identifiable fields had been removed or encrypted. The ethics and confidentiality committee of the National Information Governance Board confirmed that data could be linked in this way without explicit patient consent.

Variable definitions

For intervention patients, study endpoints were calculated over 12 months after the date of enrolment into Birmingham OwnHealth. For matched controls, we used the date of enrolment for the corresponding intervention patient. Variables were therefore calculated over the same period for matched pairs as for intervention patients.

Analysis of inpatient activity was limited to “ordinary admissions” by excluding regular ward attendances, maternity events, and transfers. Admissions were then classified into either emergency or elective admissions, based on the method of admission. Bed days included stays after emergency admission only and excluded same day admissions and discharges.

Secondary care costs included inpatient and outpatient costs. They were estimated by applying a set of unit costs specific to the case mix,¹⁷ which represented the tariff amounts that providers were allowed to charge commissioners in the United Kingdom’s health service. We did not include adjustments for regional differences in care costs, to allow robust comparison of the volume of care services between intervention patients and matched controls. We only attached costs to activity covered by mandatory tariffs, which excluded locally negotiated non-tariff payments and augmented care payments associated with critical care.

Baseline variables were derived using hospital data recorded before the enrolment dates. The variables were based on those used in an established predictive model for emergency hospital admissions over 12 months.¹⁸ These variables were age band; sex; area based socioeconomic deprivation score¹⁹; health conditions; the number of long term health conditions; and previous emergency, elective, and outpatient hospital use. There were 16 health condition variables, formed from primary and secondary codes from ICD-10 (international classification of diseases, 10th revision) on inpatient data over three years. These health conditions were anaemia, angina, asthma, atrial fibrillation and flutter, cancer, cerebrovascular disease, congestive heart failure, chronic obstructive pulmonary disease, diabetes, history of falls, history of injury, hypertension, ischaemic heart disease, kidney failure, mental health conditions, and peripheral vascular disease.

In addition to these baseline variables, we estimated the risk of emergency hospital admission in the subsequent 12 months. The predictive risk models were based on the variables used in the published model¹⁸ but reweighted to reflect patterns of hospital use of Birmingham residents who had never been enrolled into the service. Models were constructed using logistic regression on a monthly basis through the enrolment period and validated on split samples. The estimated β coefficients from the validated models were then applied to intervention patients and potential controls to produce the risk scores that would be used in the matching process.

Methods to select control group

There are several methods for selecting matched control groups, but the aim is always to select, from the wider population of potential controls, a subgroup of patients that is similar to the intervention group with respect to predictive baseline variables.²⁰ The risk score was strongly predictive, so we used a calliper approach whereby the pool of potential matches for a given intervention patient was narrowed down to those patients with a similar risk score (within 20% of one standard deviation).²¹ From within this restricted set, one control was selected for the intervention patient based on the individual baseline variables using the Mahalanobis multivariate distance metric.²² One matched control was selected for each intervention patient. This was done without replacement, so that the control group consisted of distinct individuals.

The main diagnostic in matched control studies is balance, which refers to the similarity of the distribution of baseline variables between intervention and matched control groups. Formal statistical tests are not recommended in the assessment of

balance, because they depend on the size of the groups as well as their similarity.²³ Instead, we assessed balance using the standardised difference, which is the difference in means as a proportion of the pooled standard deviation.²⁴ Although the standardised difference would ideally be minimised without limit, 10% is often used as a threshold to denote meaningful imbalance.²⁵ The ultimate aim was to select a matched control group that was well balanced across all of the baseline variables, including the predictive risk score; therefore, we adapted the set of variables included in the Mahalanobis distance until we achieved the satisfactory balance.²⁶

Statistical approach

After matched groups had been constructed, we estimated the intervention effect using a difference-in-difference estimator. Thus, intervention and matched control groups were compared in terms of the change in the number of hospital admissions observed from the year before the enrolment date to the year after the enrolment date. Paired *t* tests were conducted on the change scores to reflect the matched nature of the data.²⁷

Analysis was conducted over 12 months, regardless of death. The use of national administrative data to define variables meant that we considered there was a limited amount of missing data, because patients could be tracked even if they moved out of the Birmingham area, provided that they remained within England. We did not analyse patients who could not be linked to hospital data or could not be matched to a control.

Efforts to avoid bias and sensitivity analysis

The analysis was designed to reduce the susceptibility of the study to differences between intervention and control groups. Differences in variables that are predictive of future hospital use could result in confounding and biased estimates. The matching algorithm removed differences in important predictive variables and the difference-in-difference estimator was expected to remove the effect of all confounders that are not time varying, regardless of whether or not they were observed.

We conducted sensitivity analyses to test the robustness of our findings to time varying, unobserved confounding.²⁸ Firstly, we followed the recommendation of West and colleagues²⁹ and compared the intervention and matched control groups in terms of an outcome that we did not expect to be influenced by the intervention—namely, in-hospital mortality over 12 months (although one randomised study has reported effects of telephone health coaching on mortality³⁰). Secondly, we assessed the strength of unobserved confounding that would be required to alter our findings in relation to a dichotomised version of the our primary endpoint. Specifically, we simulated a hypothetical unobserved confounder and estimated the odds ratios that would be required between this confounder and intervention status and outcome for our findings to be altered.³¹ The values thus obtained were compared with estimates of odds ratios for unobserved confounders, based on another study.⁹

Ethics approval

The National Research Ethics Service confirmed that ethical approval was not required for this work, because it involved retrospective analysis of non-identifiable data for the purposes of service evaluation.

Results

Study populations

Of 3525 patients enrolled during the study period, 3070 (87.1%) were linked uniquely to one individual in the hospital data. Of the 455 records that did not link to hospital data, 86% had missing or incomplete personal linkage data in the service's operational system. Of 3070 patients linked to hospital data, 2703 (88.0%) patients had a history of inpatient or outpatient hospital use. Matched controls were found for 2698 (99.8%) of these patients. Data from the Birmingham OwnHealth service's operational system showed that the median duration of enrolment for the included intervention patients was 776 days (25.5 months; fig 1¹). Telephone calls typically lasted for 15 minutes.

Predictive risk models were fitted for each of the 33 months spanning participant enrolment and were validated in separate samples. The median positive predictive value was 57.0% (range 55.6% to 58.6%) and the median sensitivity was 4.3% (3.9% to 4.7%), calculated using a risk threshold of 0.5. The area under the receiver operating characteristic curve had a median value of 0.698 (range 0.688 to 0.701).

In relation to diagnoses of health conditions and other baseline variables, intervention patients differed markedly from the general population of the control areas (table 1²). For example, intervention patients had 1.2 chronic health conditions on average, compared with 0.3 conditions for the general population. However, after matching, matched controls and intervention patients had similar characteristics. For example, both groups had 1.2 chronic health conditions on average, mean age of 65.5 years, mean predictive risk score of 0.17, similar prevalence of health conditions, and similar previous hospital use (table 1, fig 2³). Standardised differences were much lower than the 10% threshold, apart from diagnoses for angina (11.1%) and mental health conditions (15.0%).

Comparing hospital use and costs

Intervention patients had more emergency admissions in the year after enrolment than in the year before enrolment (0.38 v 0.31 admissions per head). A smaller increase was observed for matched controls (table 2⁴, fig 3⁵). Comparing the two groups, emergency admissions increased by 0.05 per head more among intervention patients than among matched controls (95% confidence interval 0.00 to 0.09, $P=0.046$; table 3⁶), which was a relative increase of 13.6% (0.2% to 27.1%).

Differences between groups in hospital bed days and elective admissions were not significant (table 3). Outpatient attendances rose by 0.37 attendances per head more among intervention patients than among matched controls (95% confidence interval 0.16 to 0.58, $P<0.001$), due to a fall among controls. Overall secondary care costs increased by £175 (£203; \$268) per head more among intervention patients than among controls (£22 to £328, $P=0.025$).

Sensitivity analysis for unobserved confounding

Within 12 months of the intervention, 2.3% of patients in both the intervention ($n=63$) and matched control ($n=63$) groups died in hospital. Sensitivity analysis simulated a hypothetical unobserved confounding variable and showed that, for the apparent increase in emergency admissions to be reversed, such a variable would need to be strongly associated with both intervention status and outcome, with odds ratios greater than 2.8. By comparison, insulin treatment (which is one variable

we did not observe) had an odds ratio of 1.6 with intervention status.⁹

Discussion

Statement of findings

Telephone health coaching aims to support patients in managing their long term health conditions. The hope is that, by promoting healthy behaviours and by providing a means to identify problems before they become critical, telephone health coaching can help prevent crises that lead to hospital admissions. We compared a large sample of people receiving telephone health coaching in England to a well balanced, retrospectively matched control group using person level data. Rather than see a reduction in hospital activity in the study group, we found that emergency admissions increased at a faster rate among intervention patients than matched controls, as did outpatient attendances and secondary care costs. Therefore, there was no evidence of reductions in hospital admissions, and no savings were detected from which to offset the cost of the intervention.

Strengths and weaknesses

We were able to study a large number of intervention patients with a high rate of data linkage (87%). Imperfect linkage was mainly due to imperfect recording of individual identifiers on the service's operational system, because most records that did not link had missing or incomplete personal linkage data. On the assumption that recording omissions happened at random, our sample was an unbiased sample from the population receiving the intervention. Although the analysis then focussed on patients with previous hospital use, this variable is where the scope for savings was highest.

The use of administrative data meant that data were available for a high proportion of patients, and avoided problems of under-reporting by patients about how many services were used.³² However, the quality of data was not directly under our control. Potential problems with administrative data included limited insight into the quality and appropriateness of care,³³ and observational intensity bias if coding practices varied between geographic areas.³⁴

We obtained data for more patients than what our sample size calculation suggested was needed (2698 v 2035). Therefore, although we originally envisaged that we would only be able to detect differences in emergency admission rates of 15% or higher, the 13.6% increase detected was statistically significant, and was unlikely to be the result of chance ($P=0.046$). A 13.6% increase in emergency admissions is substantial for the health service, and much more than the general increase in age standardised rates of admission of 2.5% a year.³⁵

The main risk to validity in this observational study was that, although intervention and matched control groups were similar in terms of an established set of predictors of future hospital use, they could have differed in ways that we could not observe (that is, there may have been unobserved confounding).

Typically, only a small proportion of eligible patients receive complex interventions out of the hospital.³⁶ Birmingham OwnHealth was a relatively established service, and at least 80% of the local general practices had participating patients. Nevertheless, there are around 9000 patients in the area with uncontrolled diabetes,³⁷ for example, while only around 3000 patients received the intervention in the time period chosen.

We sought to minimise unobserved confounding by careful selection of the pool of potential controls, matching on previous outcomes and difference-in-difference estimation. The eligibility

criteria for the service included clinical variables such as HbA_{1c}. These variables were not recorded in our dataset. However, we ensured that the prevalence of health conditions was similar between the two groups, as were variables that are correlated with clinical indicators, such as hospital use.³⁸ Sensitivity analysis showed that, although the increase in emergency admissions could conceivably have been caused by unobserved confounding, it is unlikely that we missed a reduction. To have missed such a reduction, the amount of unobserved confounding would have had to be greater than is realistic for clinical variables. Further, it is reassuring that no differences were observed in in-hospital mortality between the two groups. For example, if disease control had been worse among intervention patients, more deaths might have been expected.

Observational study designs have some advantages over randomised controlled trials. This study looked at a population that was selected to participate in telephone health coaching in routine practice. By contrast, randomised controlled trials may have poor generalisability when the patients in the trial differ from those who would receive the intervention routinely. Such differences might occur because of the selection of healthcare settings or practitioners for a trial, the choice of eligibility criteria for a trial, certain individuals preferring not to participate, or study design.³⁹

This study investigated effects on hospital use and associated costs for people enrolled between 2006 and 2008. The effect of the service might have changed over time, because the eligibility criteria were later broadened to include chronic kidney disease, stroke, and transient ischaemic attack from 2009; and hypertension and older patients at risk from 2010. Although not the focus of this study, Birmingham OwnHealth might have affected the use of primary care or health related quality of life. Because the median duration of enrolment was several years, effects could have been over longer time periods than those analysed in this study. Other work has found improvements in clinical metrics among participants with poorly controlled diabetes,⁹ as well as high levels of patient satisfaction.⁸

Comparison with other studies

Previous studies of the effect of telephone health coaching on service use have generally been encouraging. Four of nine studies identified by a systematic review found evidence of an effect on health service use, although sample sizes were typically small.² A more recent, large randomised controlled trial found reductions in hospital admissions and expenditures.⁴ Before the current study, the largest observational study of telephone health coaching included 874 Medicaid members, and found no effect.⁶ The current study supports their findings on a larger sample drawn from England (n=2698).

One possible explanation for the apparently contradictory nature of the findings from these studies might be subtle differences in the design of the interventions. Aspects of intervention design such as the frequency of telephone calls vary widely between studies,² although the profile of telephone calls in the current study (usually monthly) was not out of line. The largest randomised controlled trial⁴ included decision making for preference sensitive conditions, while three of the four effective interventions identified by the systematic review involved telemonitoring of vital signs in addition to health coaching. Telemonitoring could be effective at reducing hospital admissions even when combined with automated motivational messages and symptom questions rather than health coaching.¹⁵ This study adds weight to the conclusions suggested by the Medicaid studies^{6,7} that health coaching is not effective at

reducing hospital use by itself. Further, although potentially explained by unobserved confounding, we found evidence that the intervention in Birmingham increased emergency admissions. Previous evaluations of other complex interventions out of hospital have also found indications of increases.¹⁴ One possibility is that increases occur as a result of greater observation. Indeed, in other settings, more intense observation and greater use of diagnostic tests have been found to correlate with the number of medical interventions made.⁴⁰

Discrepancies between the findings of different studies could also be due to study settings, with both the Medicaid studies and the current study relating to a publicly insured population. Targeting the intervention using the outputs from a predictive risk model may increase effectiveness, as could better integration with existing primary and secondary care services. Finally, the evaluation method could affect results. Confounding is possible in observational studies, although our study design attempted to limit this threat to validity as much as possible.

Conclusions

We conclude that Birmingham OwnHealth did not lead to the anticipated levels of reductions in hospital admissions or associated costs. Based on a systematic review and subsequent studies, including the present study, standard telephone health coaching seems unlikely to lead to reductions in hospital use, without the addition of other elements such as telemonitoring, shared decision making for preference sensitive conditions, or predictive modelling. More care coordination might also be needed. Unless health coaches have established relationships with other clinical staff, new interventions could prove to be additions to existing patterns of service use, rather than create efficiencies.

The study serves as a warning that efficacy as demonstrated by randomised controlled trials might not imply effectiveness in routine practice.⁴¹ Because administrative datasets are regularly updated, the methods used in the present study may be useful to monitor new services to ensure that benefits are achieved.

We thank staff in Birmingham East and North Primary Care Trust who organised data for scheme participants; John Grayland for his ongoing support; NHS Information for health and social care for providing invaluable support and acting as a trusted third party for the linkage to national hospital data; and Sally Inglis, Susannah McClean, and Doug Altman for their comments on a previous version of this manuscript. The data analysis for this paper was generated using the SAS software, version 9.3 (SAS Institute). SAS and all other SAS Institute product or service names are registered trademarks or trademarks of SAS Institute. Contributors: AS, IB, and MB designed the study. In addition, AS led the analysis and prepared the draft manuscript, ST liaised with the scheme about access to participant data, and IB derived unit costs for HES data. All authors reviewed the manuscript. AS was the study guarantor.

Funding: This study was funded by the Department of Health in England, which reviewed the study protocol as part of the application for funding and agreed to publication. The views expressed are those of the authors and not of the Department of Health and does not constitute any form of assurance, legal opinion or advice. The organisations at which the authors are based shall have no liability to any third party in respect to the contents of this article.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: funding from the Department of Health for the submitted work; AS, IB, and MB have a range of current or pending research grants on related topics from funding bodies including the National Institute for Health Research,

What is already known on this topic

Telephone health coaching provides support and encouragement to patients to manage long term health conditions
It is hoped that hospital admissions will be prevented as a result, creating efficiency gains for healthcare systems
However, the current evidence base is unclear; many studies have been small and interventions are heterogeneous

What this study adds

This study adds weight to the existing view that health coaching by itself is not effective at reducing hospital use over 12 months
Coaching could be coupled with other interventions such as shared decision making or telemonitoring, and the context in which interventions are delivered might also be crucial
Efficacy of new services as demonstrated by randomised controlled trials might not imply effectiveness in routine practice

Technology Strategy Board, and NHS trusts; ST, as an Ernst & Young employee, has declared that Ernst & Young is a consulting firm which may at times undertake consultancy work relevant to the commissioning and provision of community based care.

Ethical approval: The ethics and confidentiality committee of the National Information Governance Board confirmed that data linkage (as described in the methods) was possible without explicit patient consent. The National Research Ethics Service confirmed that ethical approval was not required for this work, because it involved retrospective analysis of non-identifiable data for the purposes of service evaluation.

Data sharing: No additional data available.

- McLean S, Protti D, Sheik A. Telehealthcare for long term conditions. *BMJ* 2011;342:d120.
- Hutchison AJ, Breckon JD. A review of telephone coaching services for people with long-term conditions. *J Telemed Telecare* 2011;17:451-8.
- Birmingham East and North NHS, NHS Direct, and Pfizer Health Solutions. OwnHealth Birmingham: successes and learning from the first year. September 2007. www.ehiprimarycare.com/img/document_library/0282/BirminghamOwnHealth_-_successes_and_learning_from_the_first_year.pdf.
- Wennberg DE, Marr A, Lang L, O'Malley S, Bennett G. A randomised trial of a telephone care-management strategy. *N Engl J Med* 2010;363:1245-55.
- Motheral BR. Telephone-based disease management: why it does not save money. *Am J Manag Care* 2011;17:e10-6.
- Lin WC, Chien HL, Willis G, O'Connell E, Rennie KS, Bottella HM, et al. The effect of a telephone-based health coaching disease management program on Medicaid members with chronic conditions. *Medical Care* 2012;50:91-8.
- Kim SE, Michalopoulos C, Kwong RM, Warren A, Manno MS. Telephone care management's effectiveness in coordinating care for Medicaid beneficiaries in managed care: a randomized controlled study. *Health Serv Res* 2013; doi:10.1111/1475-6773.12060.
- Azarmina P, Lewis J. Patient satisfaction with a nurse-led, telephone-based disease management service in Birmingham. *J Telemed Telecare* 2007;13(suppl 1):3-4.
- Jordan RE, Lancashire RJ, Aday P. An evaluation of Birmingham Own Health telephone care management service among patients with poorly controlled diabetes. A retrospective comparison with the General Practice Research Database. *BMC Public Health* 2011;11:707.
- NHS Direct. Chief executive's report: agenda item 6. 19 December 2011. www.nhsdirect.nhs.uk/About/CorporateInformation/MinutesOfMeetings/BoardMeetingPapers2011/~/_media/Files/BoardPapers/December2011/Item6_CEOReport.ashx.
- Steventon A, Bardsley B, Billings J, Georgiou T, Lewis GH. The role of matched controls in building an evidence base for hospital-avoidance schemes: a retrospective evaluation. *Health Serv Res* 2012;47:1679-98.
- Roland M, Dusheiko M, Gravelle H, Parker S. Follow up of people aged 65 and over with a history of emergency admissions: analysis of routine admission data. *BMJ* 2005;330:289-92.
- Office for National Statistics. National Statistics 2001 Area Classifications. www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/index/index.html.
- Gravelle H, Dusheiko M, Sheaff R, Sargent P, Boaden R, Pickard S, et al. Impact of case management (Evercare) on frail elderly patients: controlled before and after analysis of quantitative outcome data. *BMJ* 2006;390:20.413310.55.
- Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, et al. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *BMJ* 2012;344:e3874.
- Health and Social Care Information Centre. Hospital episode statistics. 2012. www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937.
- Department of Health. Payment by results: guidance and tariff for 2008-09. 2007.
- Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ* 2006;333:327.

- Communities and Local Government. The English indices of deprivation 2007. 2008.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
- Hansen BB. The prognostic analogue of the propensity score. *Biometrika* 2008;95:481-8.
- Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 1985;39:33-8.
- Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. *J R Statist Soc A* 2008;171:481-502.
- Austin P. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008;27:2037-49.
- Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol* 2001;54:387-98.
- Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal* 2007;15:199-236.
- Austin PC. Discussion of 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' [rejoinder]. *Stat Med* 2008;27:2066-9.
- Greenwood RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. *J Clin Epidemiol* 2009;52:228.
- West S.G., N. Duan, W. Pequegnat, et al. 2008. Alternatives to the randomised controlled trial. *Am J Public Health* 2008;98:1359-66.
- Alkema GE, Wilber KH, Shannon GR, Allen D. Reduced mortality: the unexpected impact of a telephone-based care management intervention for older adults in managed care. *Health Serv Res* 2007;42:1632-50.
- Higashi T, Shekelle P, Adams J, Kamberg, K, Roth C, Solomon D, et al. Quality of care is associated with survival in vulnerable older patients. *Ann Intern Med* 2005;143:274-81.
- Richards SH, Coast J, Peters TJ. Patient-reported use of health service resources compared with information from health providers. *Health Soc Care Community* 2003;11:510-8.
- Roos LL, Mustard CA, Nicol JP, McLerran DF, Malenka DJ, Young TK, et al. Registries and administrative data: organization and accuracy. *Med Care* 1993;31:201-12.
- Wennberg JE, Staiger DO, Sharp SM, Gottlieb DJ, Bevan G, McPherson K, et al. Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims. *BMJ* 2013;346:f549.
- Bardsley M, Blunt I, Davies S, Dixon J. Is secondary preventive care improving? Observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care. *BMJ Open* 2013;3:e002007.
- Subramanian U, Hopp F, Lowery J, Woodbridge P, Smith D. Research in home-care telemedicine: challenges in patient recruitment. *Telemed J E Health* 2004;10:155-61.
- Information Centre for health and social care. Quality and outcomes framework, achievement, prevalence and exceptions data, 2011/12. 2012. www.gpcwm.org.uk/wp-content/uploads/file/QOF/QOF_Achievement_prevalence%20and%20exceptions%20data%202011_12%2030%20Oct%2012.pdf.
- Govan L, Wu O, Briggs A, Colhoun HM, Fischbacher CM, Leese GP, et al. Achieved levels of HbA1c and likelihood of hospital admission in people with type 1 diabetes in the Scottish population: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetes Care* 2011;34:1992-7.
- Rothwell PM. Treating individuals 1: external validity of randomised controlled trials: "To whom do the results of this trial apply?" *Lancet* 2005;365:82-93.
- Lucas FL, Siewers AE, Malenka DJ, Wennberg DE. Diagnostic-therapeutic cascade revisited: coronary angiography, coronary artery bypass graft surgery, and percutaneous coronary intervention in the modern era. *Circulation* 2008;118:2797-802.
- Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ* 1999;319:652-3.

Accepted: 9 July 2013

Cite this as: *BMJ* 2013;347:f4585

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

Tables

Table 1 | Differences in demographic, health or healthcare characteristics of study groups before and after matching. Data are proportion (%) of individuals or mean (standard deviation) unless otherwise stated

	Potential controls* (n=969 677)	Intervention participants (n=2698)	Matched controls (n=2698)	Standardised difference (variance ratio)	
				Before matching	After matching
Age	41.6 (25.3)†	65.5 (13.4)	65.5 (13.5)	18.0 (0.28)	0.2 (0.98)
Female (%)	54.7 (n=522 049)‡	47.7 (n=1288)	47.7 (n=1288)	14.0	0.0
Socioeconomic deprivation score†	34.4 (17.5)	37.2 (18.9)	36.8 (19.3)	14.8 (1.21)	1.8 (1.03)
Anaemia (%)	1.7 (n=16 797)	3.2 (n=86)	3.7 (n=99)	9.4	2.6
Angina (%)	2.3 (n=22 092)	15.4 (n=416)	11.6 (n=314)	47.6	11.1
Asthma (%)	3.8 (n=36 971)	5.7 (n=154)	6.9 (n=186)	8.9	4.9
Atrial fibrillation and flutter (%)	2.3 (n=22 370)	6.1 (n=165)	6.5 (n=176)	19.0	1.7
Cancer (%)	5.3 (n=51 530)	6.1 (n=165)	8.0 (n=216)	3.5	7.4
Cerebrovascular disease (%)	1.8 (n=17 807)	4.2 (n=112)	4.1 (n=111)	13.6	0.2
Congestive heart failure (%)	1.6 (n=15 055)	5.7 (n=154)	5.5 (n=149)	22.4	0.8
Chronic obstructive pulmonary disease (%)	1.9 (n=18 217)	5.6 (n=152)	5.5 (n=149)	19.8	0.5
Diabetes (%)	4.0 (n=38 805)	29.6 (n=799)	29.3 (n=791)	72.9	0.7
History of falls (%)	2.3 (n=22 549)	3.2 (n=85)	2.4 (n=66)	5.1	4.3
History of injury	7.6 (n=73 450)	7.4 (n=201)	7.9 (n=213)	0.5	1.7
Hypertension (%)	8.7 (n=84 092)	28.4 (n=767)	31.2 (n=842)	52.6	6.1
Ischaemic heart disease (%)	4.0 (n=38 799)	19.9 (n=537)	19.7 (n=532)	50.6	0.5
Kidney failure (%)	1.1 (n=10 996)	2.9 (n=78)	2.9 (n=77)	12.5	0.2
Mental health conditions (%)	3.0 (n=28 910)	2.4 (n=64)	5.2 (n=141)	3.8	15.0
Peripheral vascular disease (%)	2.1 (n=20 735)	4.1 (n=111)	5.0 (n=134)	11.4	4.1
Number of long term conditions	0.3 (0.8)	1.2 (1.5)	1.2 (1.5)	74.6 (3.41)	0.7 (1.02)
Predictive risk score	0.10 (0.08)	0.17 (0.12)	0.17 (0.2)	67.3 (2.15)	0.0 (1.00)
Emergency admissions (previous year)	0.2 (0.6)	0.3 (0.7)	0.3 (0.7)	19.7 (1.53)	2.1 (1.14)
Emergency admissions (previous month)	0.01 (0.13)	0.03 (0.21)	0.03 (0.17)	9.7 (2.37)	3.2 (1.49)
Elective admissions (previous year)	0.3 (1.4)	0.4 (1.1)	0.3 (0.9)	4.8 (0.56)	2.9 (1.44)
Elective admissions (previous month)	0.03 (0.21)	0.03 (0.19)	0.03 (0.19)	2.2 (0.81)	0.8 (1.02)
Outpatient attendances (previous year)	1.6 (3.0)	3.9 (5.0)	3.6 (4.5)	55.5 (2.66)	6.1 (1.24)
Outpatient attendances (previous month)	0.15 (0.54)	0.36 (0.75)	0.33 (0.72)	32.8 (1.88)	3.9 (1.07)
Emergency bed days (previous year)	1.06 (7.64)	1.80 (8.02)	1.57 (6.16)	9.5 (1.10)	3.3 (1.70)
Emergency bed days (previous year, trimmed to 30 days)	0.77 (3.74)	1.44 (4.79)	1.38 (4.58)	15.6 (1.64)	1.3 (1.09)

*Residents of control areas with previous hospital use.

†For complete cases (n=968 659).

‡For complete cases (n= 954 282).

†Taken from the Index of Multiple Deprivation 2007.¹⁹

Table 2| Estimated intervention effects for secondary care use

	Participants (n=2698)			Matched controls (n=2698)		
	Year before enrolment	Year after enrolment	Difference	Year before enrolment	Year after enrolment	Difference
Emergency admissions	0.31 (0.74)	0.38 (0.93)	0.07* (0.96)	0.29 (0.69)	0.32 (0.88)	0.03 (0.95)
Bed days (after emergency admission)	1.80 (8.02)	2.90 (11.85)	1.09* (12.86)	1.57 (6.16)	2.57 (10.79)	1.00* (11.29)
Elective admissions	0.37 (1.05)	0.39 (1.10)	0.02 (1.20)	0.34 (0.87)	0.38 (1.11)	0.04 (1.25)
Outpatient attendances	3.90 (4.96)	3.96 (5.13)	0.06 (4.29)	3.61 (4.45)	3.30 (4.91)	-0.30* (4.28)
Secondary care costs (£)	1388 (2683)	1650 (3228)	261* (3534)	1292 (2436)	1379 (3054)	86 (3355)

Data are mean number or cost per head (standard deviation). *Difference=P<0.05.

Table 3| Difference-in-difference estimate of intervention effect (per head)

	Mean (standard deviation)	95% confidence interval	P
Emergency admissions	0.05 (1.20)	0.00 to 0.09	0.046
Bed days (after emergency admission)	0.09 (15.87)	−0.50 to 0.69	0.757
Elective admissions	−0.02 (1.48)	−0.07 to 0.04	0.549
Outpatient attendances	0.37 (5.50)	0.16 to 0.58	<0.001
Secondary care costs (£)	175 (4592)	22 to 328	0.025

Figures

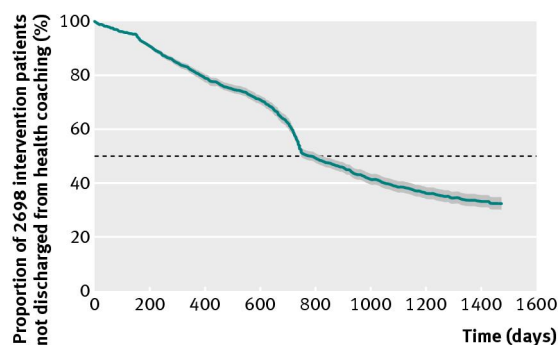


Fig 1 Length of time spent enrolled in the Birmingham OwnHealth service. Solid line=best estimate; shaded area=95% confidence interval

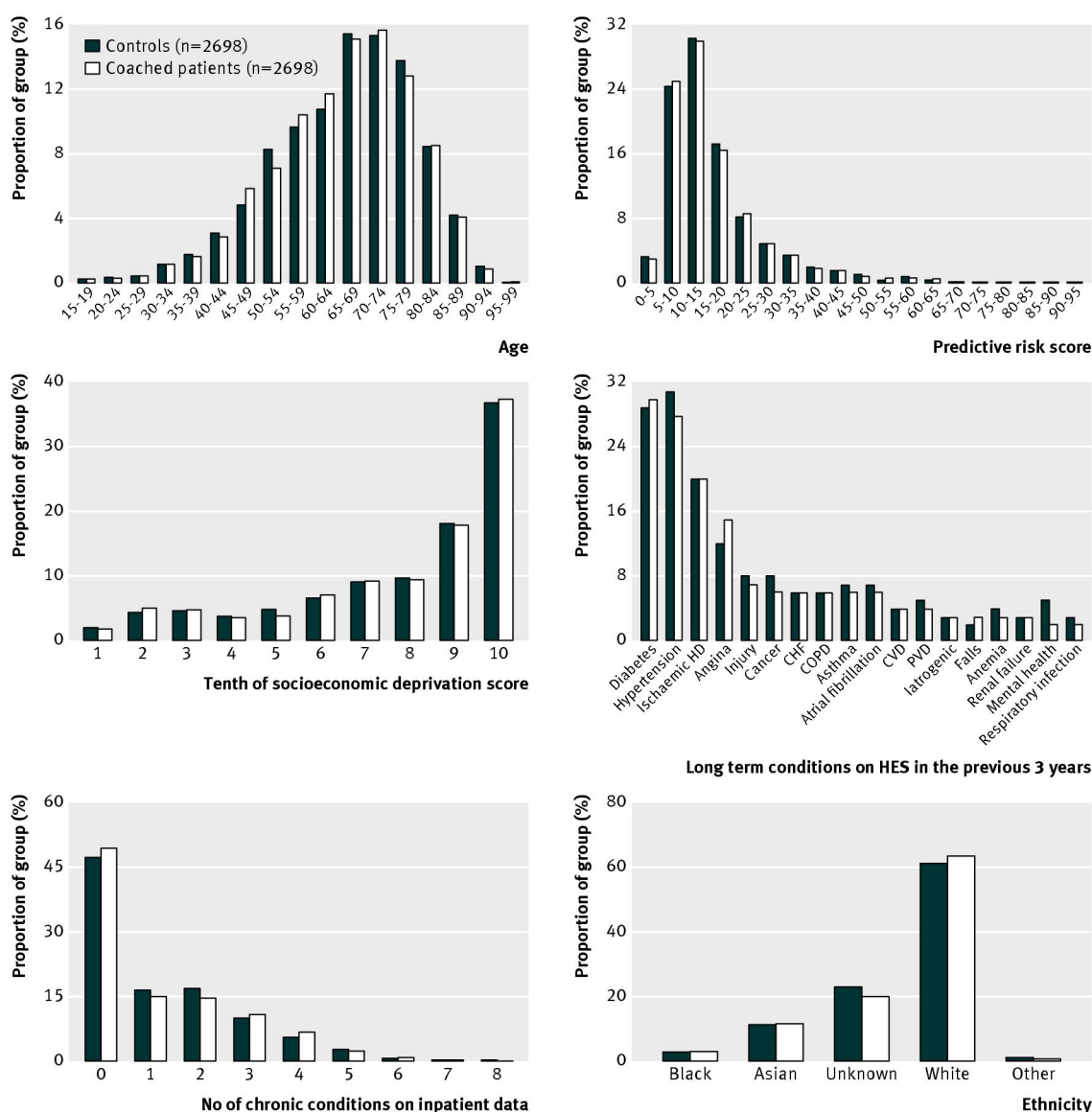


Fig 2 Differences between 2698 coached patients and 2698 matched controls at the start of intervention. HD=heart disease; CHF=congestive heart failure; COPD=chronic obstructive pulmonary disease; CVD=cerebrovascular disease; PVD=peripheral vascular disease. *Score of 10=most deprived; tenths were defined on the basis of national data for the Index of Multiple Deprivation 2007¹⁹

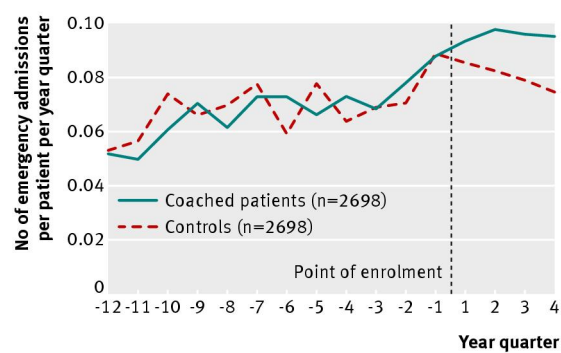


Fig 3 Comparison of rate of emergency admissions