Patterns of Benign Gynaecology Care in English NHS Hospital Trusts

2015/16

Royal College of Obstetricians & Gynaecologists

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Table of Contents

Abbreviations **3**

Glossary of terms **5**

Acknowledgements **7**

Foreword **8**

Executive summary **9**

1. Introduction **13**

1.1. Purpose of the report 13

2. Indicator Development Process **15**

2.1. Identification 15

2.2. Development and Evaluation 17

3. Data source **20**

3.1. Data source: Hospital Episode Statistics 20

4. Methodology for deriving indicators **20**

4.1. Cohort selection 20

4.2 Analysis and case mix adjustment 21

Relationships between indicators 21

4.3 Presentation of data using funnel plots 22

5. Results **24**

5.1. Hysterectomy Indicators 24

Construction of the hysterectomy type indicators 24

Length of stay following hysterectomy 25

Emergency readmission within 30 days of hysterectomy 31

5.2. Mid-urethral Mesh Sling Indicators 33

Length of stay following MUS insertions (TVT/TOT) 33

Emergency readmission within 30 days of MUS insertion 34

Reoperation within 2 years of a mid-urethral mesh sling insertion 35

5.3. Prolapse Repair Indicators 36

Length of stay following pelvic organ prolapse (POP) repair 36

Emergency readmission within 30 days of POP repair 37

Reoperation within 2 years of a POP repair procedure 38

5.4. Endometrial Ablation Indicators 39
Length of stay following endometrial ablation 39
Emergency readmission within 30 days of endometrial ablation 40
5.5 Summary of results 41
6. Discussion 42
Hysterectomy type 42
Length of stay 42
Emergency readmission 43
Reoperation following MUS insertion or POP repair 44
7. Conclusions 52
Appendices 54
Appendix 1. Additional detail on the data and analytical methodology 54
Appendix 2. Sources of indicators identified by the review 57
Appendix 3. Indicator Consensus Group Membership 58
References 60

Table of Tables

Table 1: Evaluation of potential benign gynaecology indicators 18
Table 2: Demographic characteristics of women in the HES benign gynaecology sample 21
Table 3: characteristics of women undergoing hysterectomy by condition 24
Table 4: Length of stay by type of hysterectomy 26
Table 5: Summary of findings 41
Table 6: Number of procedures required to detect a rate double the national mean 47

Table of Figures

Figure 1: How to develop relevant, rigorous and robust clinical indicators 16
Figure 2: How to interpret a funnel plot 22

Abbreviations

AH	Abdominal hysterectomy
EA	Endometrial ablation
H	Hysterectomy
HES	Hospital Episode Statistics
HESID	Hospital Episode Statistics Identification Number
HMB	Heavy menstrual bleeding
ICD-10	International Classification of Diseases, 10th edition
LH	Laparoscopic hysterectomy
LSC	Lindsay Stewart Centre for Audit and Clinical Informatics
LSHTM	London School of Hygiene and Tropical Medicine
MUS	Mid-urethral mesh sling
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
OPCS	Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures, 4th revision
POP	Pelvic organ prolapse
RCOG	Royal College of Obstetricians and Gynaecologists
SUI	Stress urinary incontinence
TOT	Trans-obturator insertion of mid-urethral mesh tape
TVT	Retropubic insertion of mid-urethral mesh tape
VH	Vaginal hysterectomy

Glossary of terms

Case mix

Clinical and demographic characteristics of patients affect both the demands placed on the service and the outcomes of care. Case mix is a term used to refer to how similar the patient groups are across organisations and it needs be taken into account when organisations are compared.

Fairness

The extent to which an indicator used for comparative purposes takes into account differences in case mix between trusts.

Hospital Episode Statistics (HES)

A data 'warehouse' that includes records of all inpatient admissions, day cases, outpatient clinic and accident-and-emergency department visits in English NHS hospitals, with the data being extracted from local patient administration systems. It is managed by NHS Digital.

Indicator

A statistic that can be used to describe levels of performance of a health service provider that, in turn, can help identify possible problems and/or opportunities for improvement of the service.

Multiparous

The medical term used to describe a woman who has given birth before.

Outcome indicator

A type of indicator which measures the outcome of care received. Outcome indicators can be difficult to interpret as differences between organisations in terms of outcome indicators do not necessarily reflect differences in the quality of care.

Process indicator

A type of indicator which measures a process of care (what was done to whom, and when). Process indicators are valid measures of quality if they are based on strong evidence for the effectiveness of a particular treatment or intervention: the more patients - without contraindications – who receive a proven treatment or intervention, the better.

Random variation

A statistical term that refers to the tendency for the estimated value of a parameter to deviate randomly from the true value of that parameter ('the play of chance'). In general, the larger the sample size, the lower the impact of random variation on the estimate of a parameter. As random variation decreases, precision increases.

Risk adjustment

When presenting figures for individual trusts, indicators must take into how similar the patient groups are ('case mix') that are being treated in the trusts. Risk adjustment is a statistical technique that controls for differences in the clinical and demographic characteristics of patients, which are outside the control of care providers. It is essential for fair and meaningful comparisons across trusts.

Routine data

Data routinely collected by hospitals, often for administrative purposes. The data is primarily intended for health service planning and guiding the reimbursement of health care expenses, but can also be used to study patterns of care. HES is an example of a routine database.

Statistical power

The ability of a statistical test to detect a meaningful effect. It depends on the sample size (the number of patients included in a particular comparison), the statistical significance level of the test, and the size of effect defined as meaningful.

Validity

The extent to which an indicator reflects quality of care. An indicator is valid if differences in the indicator reflect differences in the quality of care, and if it is clear which end of the indicator spectrum represents high quality.

Acknowledgements

We would like to thank the Clinical Effectiveness Unit, the collaborative unit of the London School of Hygiene and Tropical Medicine and the Royal College of Surgeons of England for their support with the analysis for this report.

We are indebted to the members of the Indicator Consensus Group (see Appendix 3 for membership) who guided the selection and technical specification of the indicators described in this report, as well as to Justin Clark, who volunteered to write the expert opinion piece interpreting the report's findings.

The RCOG Clinical Indicators Programme is part of the Lindsay Stewart Centre for Audit and Clinical Informatics (LSC). This work was funded in part by a generous legacy donation made by Dr Lindsay Stewart OBE CA FRCOG(hon) FRCSEd FRCSI to support research and development activities within the RCOG. We are grateful for the oversight of the LSC committee, and for the helpful input on the project, and comments on this report, from members of the committee, and Professor Steve Thornton, chair of the LSC committee. We would also like to thank the membership of the various other RCOG Committees that provided feedback on draft versions of this report, including the RCOG Council.

Foreword

The RCOG Clinical Indicators Programme, and the National Maternity and Perinatal Audit, carried out by the RCOG in partnership with the RCM, RCPCH and LSHTM, have led the way in developing robust, clinically relevant indicators to facilitate monitoring of the clinical effectiveness and efficiency of maternity care in England, Scotland and Wales.

Benign gynaecology services make a fundamental contribution to women's health, but there is no standard set of performance indicators to monitor and drive quality improvement for these services. Now, expanding the RCOG's Clinical Indicators Programme, we are pleased to present this report on the first stage of our development work in benign gynaecology. Our aim is to demonstrate the feasibility of deriving national performance indicators for benign gynaecology care in England from routinely-collected data that are clinically meaningful, methodologically rigorous and technically robust.

By combining detailed analyses and clinical commentaries, the RCOG's Clinical Indicators Programme aims to lead by example and ensure the validity and fairness of the indicators we propose. In this way, we also aim to support other national initiatives currently active in the field of quality improvement in gynaecology. In the current era of assessment and accountability, this report, and other quality improvement initiatives, present a strong opportunity to improve gynaecology services and we welcome the commitments and efforts already made by these initiatives and NHS hospital trusts to address quality and safety.

In this report, we present seven indicators focussing on inpatient care, emergency readmission and longer-term surgical outcomes that can be derived from routinely-collected data. However, to create a more balanced picture of the quality of benign gynaecology care in the future, measures of women's experience, their health and quality of life before and after treatment would be particularly valuable.

The next aim of the RCOG's Clinical Indicators Programme will be to develop national indicators for gynaecological cancer care. The first stage of this work has been initiated.

We recognise that NHS hospital trusts are already actively engaged in efforts to reduce unwarranted variation and improve care. We hope that this report will act as an enabler to those efforts, empowering trusts to identify both causes of variation at a local level and opportunities to improve care. The RCOG is committed to supporting gynaecology services by producing meaningful performance indicators in a timely manner.

I hope this will be the first of many such reports.

Eddie Morris
Vice President (Clinical Quality), Royal College of Obstetricians and Gynaecologists

Executive summary

Challenges for gynaecology services

Benign gynaecology services make a fundamental contribution to comprehensive health care systems. Every year in England an estimated 50,000 women are referred to NHS secondary care for heavy menstrual bleeding alone, with approximately 28,000 of these women going on to receive surgical treatment(1, 2). Studies have found regional variations in referrals and surgery for HMB in England large enough to suggest scope for improvements in how HMB is managed(3, 4). Urinary incontinence meanwhile affects 30% to 40% of adult women in England with concerns about potential under-provision of surgical care, including inequitable provision to certain groups(5-8). Hysterectomy, the most frequently performed major gynaecological surgical procedure, is one of the most commonly performed surgeries worldwide(9), and the majority of hysterectomies are conducted for benign indications(10).

Performance measurement in benign gynaecology care

The growth of evidence on variation in gynaecology care has coincided with increasing demands on hospitals and trusts to publish information on the quality of the care they provide(11, 12). Such information aims to fulfil various roles: informing policy making at local and national levels, supporting clinicians and providers to improve care through comparative benchmarking, identifying unexpected levels of performance and protecting public safety, and providing information to facilitate choice of care provider.

However, there is little consensus about what information should be published. Many performance indicators have been proposed. *Outcome indicators* are of intrinsic interest and are crucial in the assessment of patient safety, but pose problems for monitoring the quality of care for benign gynaecology conditions, because many serious adverse outcomes are relatively rare. For example, intra- and post-operative mortality may act as a sentinel indicator to trigger investigations. However, these outcomes are extremely rare for benign gynaecology surgical procedures, meaning that the signal-to-noise ratio is too low for these outcomes to be used for quality improvement. Important differences in the quality of care can be 'lost' in statistics on mortality, or other rare outcomes(13). Where *process indicators* are based on strong evidence for the effectiveness of a particular treatment or intervention they provide a valuable alternative for monitoring the quality of care. However, using process indicators to measure quality in benign gynaecology care is complicated by the fact that the 'best' care often depends upon the individual context, which includes the woman's future fertility preferences as well as factors such as past obstetric history and the presence of pre-existing clinical conditions.

The RCOG Clinical Indicators Programme

The Royal College of Obstetricians and Gynaecologists has initiated a programme of work to develop valid, clinically relevant, methodologically rigorous and technically robust performance indicators for maternity and gynaecology care that improve upon the comparative information currently available. This project, carried out in collaboration with

the London School of Hygiene and Tropical Medicine (LSHTM), set out to examine the validity of potential performance indicators, and to determine how successfully these could be used to compare performance between English NHS trusts using administrative hospital data.

The RCOG Clinical Indicators Programme has already published two reports (in 2013 and 2016) describing the patterns of maternity care in England. The first of these reports published risk-adjusted results for maternity indicators for English NHS hospitals on an unnamed basis. This first report represented a feasibility stage to define and derive national performance indicators for maternity care, so publishing the results on a named-trust basis was not deemed appropriate. The second maternity indicators report was then able to report trust-level results on a named basis. This report represents a feasibility study on the derivation of national performance indicators for benign gynaecology care in England. As in the feasibility stage of the maternity indicators project we will therefore not present results on a named-trust basis, but will share individualised trust-level results with NHS trusts to enable them to review their results and compare these against national means.

Here we present a suite of seven risk-adjusted indicators that can be derived using Hospital Episode Statistics data. Their selection and technical specification was informed by a systematic review of the international literature and guided by a national consensus panel of clinical and methodological experts. The proposed indicators were developed using a transparent approach with explicit criteria to evaluate their validity, technical specification and statistical power, and risk adjustment was conducted to enable fairer comparisons between trusts serving varied populations. This robust methodology and multidisciplinary approach is a key strength in ensuring the validity and fairness of the indicators.

In addition to the RCOG clinical indicators programme, a number of other current national initiatives share the important aim of providing healthcare providers and commissioners with meaningful, high-quality information that allows them to identify priority areas for improving care and outcomes for women. By combining detailed analyses, risk-adjustment for case-mix and clinical commentaries, the RCOG's Clinical Indicators Programme aims to lead by example and ensure the validity and fairness of the indicators we propose. In this way, we also aim to support other national initiatives currently active in the field of quality improvement in gynaecology care. This report, and other quality improvement initiatives, present a strong opportunity to improve gynaecology services and we welcome the commitments and efforts already made by these initiatives and NHS hospital trusts to address quality and safety. To maximise the impact of indicators derived from administrative data for quality improvement, a combination of peer-to-peer discussion and the use of robust methodologies to adjust for differences in populations, may help to drive clinically-led discussion and quality improvement. NHS trusts and clinicians in England can use these indicators and the accompanying clinical commentaries as a basis for local reflection on practice to identify both causes of variation and opportunities to improve care, including as part of the Getting it Right First Time programme.

We intend this to be the first of an annual account of variation in care delivered to women for benign gynaecological conditions. The indicators focus on inpatient care, emergency readmission and longer-term surgical outcomes. At present there are insufficient data available to present a complete picture of quality of care throughout the benign

gynaecology care pathway in England. Measures of women's experience and quality of life before and after treatment, as well as data from outpatient care would be particularly valuable. The availability of more clinically detailed data would enable the improvement of existing indicators through refinement of risk adjustment models, and the development of new indicators to produce a more balanced picture of the quality of benign gynaecology care. Future work towards this could include using inpatient data linked with other clinical data, and the collection of user experience measures. Until then, gynaecology service user groups may help to provide contextual information that trusts can use to inform interpretation of their results. We recognise that trusts are already actively engaged in efforts to reduce unwarranted variation: this report is intended to act as an enabler to that process.

Despite the limitations of the currently available data, the initial indicators suggest some variation in both practice and outcomes between NHS trusts in England. We caution against the over-interpretation of observed patterns of care. Variation between organisations can arise from factors other than the quality of care, including random fluctuations and differences in data quality and case mix. Conclusions about quality of care can only reasonably be drawn after differences due to these factors have been excluded. At this stage of this project, the influence of all these factors cannot be fully differentiated. Despite these limitations, these results should act as a trigger for reflection by local services, leading to improvements in terms of indicator design and the quality of benign gynaecology care.

Structure of the Report

In Section 1 we explain the rationale behind the project and the purpose of this report. Section 2 describes the data source, including its advantages and limitations for deriving performance indicators. In Sections 3 and 4 we describe how the indicators were selected and derived, including how they were risk-adjusted for case-mix variation. The results are presented in Section 5. We use funnel plots to describe variation over and above that expected by chance alone. In the interests of transparency, we give the exact definitions used to derive the indicators, and the methods used to assess data quality. Finally, in Section 6 we explore possible reasons for the observed variation in the indicators, present conclusions regarding the extent to which administrative data can be used to derive robust indicators for benign gynaecology and make recommendations for the way forward.

Key Findings and Recommendations

We recommend NHS trusts and clinicians in England use these indicators as a basis for reflection on current practice, to identify both causes of variation at a local level and opportunities to improve care. Together with researchers, work is also needed to begin to define acceptable levels of variation and to understand relationships between indicators.

We show that administrative hospital data can be used to develop indicators for benign gynaecology. We present national variation in:

- hysterectomy type
- length of stay and emergency readmission within 30 days for: hysterectomy, mid-urethral mesh sling insertions, pelvic organ prolapse repair and endometrial ablation
- reoperation within two years of a mid-urethral mesh sling insertion for urinary incontinence or pelvic organ prolapse repair

However, these indicators do not cover important aspects of care, such as women's experience, or primary or outpatient care, which are not available in inpatient hospital data.

- **Researchers should use linked data to create a more balanced indicator suite;** priority areas should be primary care (including consultation, treatment and referral data) and outpatient visits in secondary care. Audits and registries should collect patient-reported measures of clinical outcomes, quality of life & experience

More than 40% of hysterectomies for benign conditions were performed abdominally, with substantial variation (mean proportion conducted abdominally: 17-67% (lowest vs. highest decile)).

- Where technically feasible hysterectomy should be performed using a minimally invasive or vaginal route
- **Trusts should review whether they provide the full range of hysterectomy types** so treatment decisions are not restricted by hospital factors

There was significant variation in length of stay for procedures often considered day cases, which may reflect differences in trust-level recovery and discharge procedures, or quality of care.

- **Trusts should explore reasons for longer lengths of stay locally to identify opportunities to improve care**

Emergency readmission rates varied substantially for all procedures between trusts in the highest vs. lowest decile.

- **Trusts should locally audit their rates of, and reasons for, emergency readmission to identify opportunities to enhance care**

Evaluation of the nature and extent of prolapse is subjective.

- Reoperation rates following pelvic organ prolapse repair may be amenable to reduction with increases in **full prolapse assessments & multi-compartment repairs**

Administrative hospital data cannot capture many issues that women may have faced after surgery for prolapse repair or mid-urethral mesh sling insertion for urinary incontinence, such as pain, dyspareunia or recurrence of the original condition where these did not lead to further surgery.

- Any future national prospective registry of mesh sling insertions for urinary incontinence & pelvic organ prolapse should **include patient-reported outcomes**

One in six trusts lay outside the funnel limits for more than one indicator.

- Further work is needed to understand relationships between indicators and which are best suited to different aims (safety, assessment, quality improvement)

1. Introduction

1.1. Purpose of the report

There are increasing demands on providers to monitor and publish information on quality of care and patient outcomes. One stimulus for this is the commitment of the government of all four nations within the UK to develop a quality assurance framework for NHS services. This framework will be used to assess performance; drive forward quality improvement; increase transparency, and aid accountability in the NHS(12).

Gynaecology services are a major component of NHS hospital services. An estimated 50,000 women are referred to NHS secondary care each year for heavy menstrual bleeding (HMB), and referrals for menstrual disorders account for 20% of referrals to specialist gynaecology services. An estimated 30,000 women undergo hysterectomy for benign indications each year(14). By the age of 80, 11% of women will have undergone surgery for pelvic organ prolapse (POP) or urinary incontinence(15) at some point in their lifetime.

The safety and quality of care delivered by maternity units in the UK has attracted a high level of public interest. Less focus has been placed on benign gynaecology care, despite accounting for 733,699 inpatient episodes in 2015-16. Recent publications have highlighted concerns about the safety and potential inequity and regional variation in some areas of benign gynaecology care. Geographical variation in surgical treatment rates for HMB are sufficiently large that they suggest there is scope to improve the management of HMB in England(3). In addition, women living in more deprived areas had more severe HMB symptoms and poorer quality of life at referral to outpatient clinics, which may reflect inequitable access to, or utilisation of, secondary care(4). Specific studies using administrative data have highlighted variation across hospitals in clinical practices such as the rate of surgery for HMB(3) as well as in adverse outcomes following MUS insertion for stress urinary incontinence(16, 17).

Clinical practice and outcomes are increasingly being described using routinely collected administrative hospital data. These data are collected primarily for health service planning and guiding the reimbursement of health care expenses. However, they can also be a valuable source of data for measuring healthcare outcomes. A number of organisations are now using routine data to produce annual statistics for English NHS services. These include both public-sector organisations and commercial companies, which aim to provide information that is relevant to women, clinicians and hospital managers. Where available, routinely collected administrative hospital datasets are attractive data sources for comparing healthcare performance due to their often large sample sizes, lack of selection bias, and the relatively low costs of accessing these data compared to conducting primary data collection. However, using administrative data to develop indicators to monitor performance can present challenges including variation in data quality, limited clinical detail, and a lack of measures of structural and user experience. As a result, using administrative hospital data for performance monitoring requires a robust methodology and cautious analysis. Furthermore, many indicators currently derived from these administrative data are not easy to interpret, due to the absence of evidence for best practice for particular situations and lack of validity.

To improve the usefulness of the information being produced on NHS services the Royal College of Obstetricians and Gynaecologists (RCOG) has developed a repository of clinical data to provide information that can be used to monitor the quality of care within the specialties. This report describes the first phase of this work for benign gynaecology, namely the identification, selection, derivation and validation of a suite of indicators using currently available, administrative hospital data from England. An important part of this first stage has been the development of methods to enable fair comparisons. This work represents a first step towards the development of a balanced suite of indicators to monitor and improve quality of care.

This report describes variation in practice and outcomes across NHS benign gynaecology services in England. However, we would caution against the over-interpretation of the observed patterns. The causes of variation in health care are complex and can arise from:

1. the influence of **random fluctuations**
2. differences in **data quality** between trusts
3. differences in **case mix** between trusts
4. differences in the **quality of care** provided

Conclusions about quality of care can only be drawn after differences due to factors 1-3 have been excluded. At this stage, the influence of all these factors cannot be differentiated. In particular, for benign gynaecology, where the prevalence of some serious adverse outcomes is low in statistical terms, there is a need to consider which indicators are most appropriate for a given aim (quality of care, safety or effectiveness), and how they can be robustly derived to facilitate comparisons between trusts. We anticipate that over time, the suite of indicators will be expanded to give a more comprehensive picture of benign gynaecology care from a woman's initial contact with primary care services through outpatient to inpatient care. Key areas for development will be measures of service user experience. This report should act as a trigger for reflection upon their clinical practice by local services, and lead towards improvements in indicator design and the quality of benign gynaecology care and women's experiences.

2. Indicator Development Process

Indicators are statistics that can describe clinical performance. The information they provide can be used for identifying possible problems and opportunities for improvement, informing policymaking, comparative benchmarking, and providing information to facilitate choice of healthcare provider.

Administrative hospital data are attractive for comparing performance of hospital-trusts due to their often large sample sizes, lack of selection bias, and the relatively low costs of accessing these data compared to conducting primary data collection. However, using administrative data to develop indicators can also present challenges. These include varying data quality, limited detail on some clinical risk factors (such as BMI, smoking and alcohol consumption), and a lack of structural and user-experience measures. We adopted a transparent approach with explicit criteria to develop performance indicators for benign gynaecology services using Hospital Episode Statistics (HES), the national administrative database of the English National Health Service (NHS). HES is described in more detail in Section 3. A similar approach has been used to develop maternity care indicators using HES and includes methods to address the challenges administrative data pose(18). This approach used a three-stage process of 'identification', 'development and evaluation', and 'implementation and feedback' (Figure 1 reproduced with permission from BJOG(18)).

2.1. Identification

We began the identification stage by conducting a systematic review of indicators used to describe the clinical practices and health outcomes of benign gynaecology services internationally, in high-income countries. We searched electronic databases[i] for studies (published since 1990), which proposed, developed or used performance indicators for benign gynaecological care in high-income countries. We identified 20 different indicators (Table 1) from the literature, which were potentially derivable from HES. Precise definitions for each indicator were documented where these were given. Next, we evaluated whether these indicators met the validity, fairness and statistical power criteria and established whether their technical specification could be met using HES data (Figure 1). In doing so, the list of potential indicators was refined to a shortlist of five indicators. To reduce the potential impact of publication bias, we also sought suggestions for indicators from a national consensus panel of clinical and academic experts. This led to the development of a further two indicators, giving a total of seven indicators. The panel comprised representatives from the gynaecology profession (surgeons and specialist nurses), statisticians, women with recent experience of benign gynaecology health care, and health service researchers (Appendix 3). The final selection of indicators represents those that were both described in the published medical literature and currently possible to derive from routinely collected, available data, rather than an ideal set of benign gynaecology indicators covering all aspects of quality, from primary care through to post-operative care and longer-term outcomes.

[i] The databases searched were: Cochrane's: Central Register of Controlled Trials (CENTRAL), Database of Systematic Reviews (Cochrane Reviews), Database of Abstracts of Reviews of Effects (Other Reviews), Health Technology Assessment Database (HTA), as well as MEDLINE, Ovid and CINAHL Plus.

FIGURE 1: HOW TO DEVELOP RELEVANT, RIGOROUS AND ROBUST CLINICAL INDICATORS

2.2. Development and Evaluation

Given the challenges of using administrative hospital data it is important to rigorously evaluate candidate indicators to address these challenges as much as possible(18). This second stage entailed evaluation against four criteria: 'validity', 'statistical power', 'technical specification' and 'fairness' (Stage 2 of Figure 1). The consensus panel assessed each shortlisted indicator against each criterion.

The results of the panel's evaluation are presented in Table 1. The final selection of seven indicators is skewed towards procedure-specific, post-operative outcomes. Procedure-specific indicators give information on outcomes of patients undergoing a specific procedure. Condition-specific indicators give information on outcomes of patients with a specific-condition from the time of diagnosis (or another time-defining event in the course of the condition). Condition-specific indicators reflect the impact of all clinical specialties involved in the treatment of patients along the entire pathway. As a result, it is difficult in many situations to link condition-specific indicators to individual NHS trusts and clinicians, and often more difficult to define the patient population and timing of follow-up than with procedure-based indicators. It is also likely that the impact of case mix differences will be greater for condition-specific than for procedure-specific indicators.

To provide a broad understanding of the performance and quality of a healthcare service it is important that a suite of indicators is 'balanced', ideally including indicators relating to the structure, processes, or outcomes of the care received throughout the pathway (from primary care to outpatient and inpatient care), and including users' experiences and patient-reported outcomes(19). Indicators based on administrative data will tend to focus on process and outcome indicators as structural and user experience measures are not normally available in these datasets (although they are important for understanding many outcomes). The measurement of quality in gynaecology care is made more complicated by the fact that the 'best' care pathway may be dependent upon a range of factors, including age, parity, past obstetric history (e.g. previous caesarean section), fertility intentions, and the presence of pre-existing conditions. The clinical and demographic characteristics of patient groups at different trusts, known as case-mix, may vary substantially and can affect the demands placed on a service and the outcomes of care. It is also important to note that the selected indicators are not entirely independent of one another. For example, a longer average length of stay for a given procedure may be reflected in a lower readmission rate. These relationships must be taken into account by trusts when interpreting their results.

TABLE 1: EVALUATION OF POTENTIAL BENIGN GYNAECOLOGY INDICATORS

Indicator	Selected (✓/×)	Reason for exclusion (if applicable)
Theme: Mortality and morbidity		
Gynaecologic death/death within 30 days of surgery	×	Power issue (death following benign procedures (statistically) too rare an event to reliably detect variation). Technical specification not possible within HES (only in-hospital deaths captured)
Cardiopulmonary arrest	×	Power issue (following benign procedures (statistically) too rare an event)
Theme: Timing or patient population		
Hysterectomy among woman aged <30 years	×	Power issue (too rare an event)
Abortions performed before the 10th week	×	Technical specification not possible in HES (no information on medical management of abortion or private providers)
Haemodynamically stable ectopic pregnancy patients seen within 4 hours of A&E book in time	×	Technical specification not possible in inpatient HES (no codes to identify haemodynamically stable patients). Time to initial assessment, and to treatment, already captured in the national A&E Quality Indicators
Theme: Surgical approach		
100% of stable ectopic pregnancy patients had laparoscopic surgery	×	Technical specification not possible in HES (no codes to identify stable patients)
85% of unstable ectopic pregnancy patients had laparoscopic surgery	×	Technical specification not possible in HES (no codes to identify stable patients)
Type of management of ectopic pregnancy	×	Technical specification not possible in HES (expectant/medical management not captured in HES)
Proportion of hysterectomies conducted abdominally	✓	-
Proportion of (all/abdominal) hysterectomies conducted laparoscopically	✓	-
Theme: Peri- or post-operative morbidity		
Septic shock rate	×	Power issue (too rare an event (statistically) following benign procedures))
Venous thromboembolism rate	×	Power issue (too rare an event)
Unplanned return to theatre	×	Power issue (too rare an event)
Unplanned admission to Intensive Care after hysterectomy	×	Power issue (too rare an event). Technical specification not possible in HES.
Injury or repair of organ during a planned procedure	×	Power issue (too rare an event). Technical specification not possible in HES (not possible to distinguish from returns to theatre on the same day as the planned procedure).
Blood loss requiring >4 units transfusion	×	Technical specification not possible in HES (no codes available on volume transfused)
Proven DVT, PE or unscheduled administration of intravenous heparin	×	Technical specification not possible in HES (no codes to indicate heparin use)
Persistent trophoblastic rate <10% after conservative management of tubal ectopic pregnancy	×	Deemed a poor indicator as not clear that this indicates quality of care
Length of stay in hospital (days)	✓	-
Theme: Emergency readmission		
Emergency readmission within 30 days of surgery	✓	-
Theme: Reoperation (identified by the consensus group)		
Reoperation for stress urinary incontinence (SUI) within 2 years of SUI (MUS insertion) surgery	✓	-
Reoperation for pelvic organ prolapse (POP) within 2 years of POP surgery	✓	-

Box 1. How the indicators were developed: A summary

- A national panel of clinical and academic experts, including gynaecologists, nurses, statisticians, health services researchers and women with recent experience of English NHS benign gynaecology services, guided the selection and technical specification of the indicators.
- Twenty existing benign gynaecology indicators were identified as potentially derivable from HES and assessed according to explicit evaluation criteria: validity, fairness and statistical power. The consensus group proposed a further two.
- Seven of these indicators were identified that met the validity, fairness and statistical power criteria and whose technical specification could be met by HES data. These indicators were further developed and are presented in this report.
- Each trust's data quality was carefully assessed. Indicators were not calculated for those with missing data in key fields. Trusts performing fewer than 500 benign gynaecology procedures per year were also excluded as such small samples would increase the impact of random fluctuations on the indicators values.
- Indicators were derived for appropriate subgroups of women and risk adjusted for relevant demographic and clinical factors to allow fair comparisons to be made.
- Results are presented using funnel plots. These visual representations allow the size of each trust to be taken into account when comparing performance.

3. Data source

3.1. Data source: Hospital Episode Statistics

The Hospital Episode Statistics (HES) data 'warehouse' contains information on each episode of admitted patient care in English NHS hospital-trusts. These data are extracted from local patient administration systems as part of the Commissioning Data Set. This is submitted to NHS Digital for processing and made available for audit and research as the HES dataset(20). Each record contains data on patient demographics (such as age, sex, ethnicity and postcode), the episode of care (e.g. hospital name, date of admission and discharge) and clinical information. Diagnoses are recorded using the International Classification of Diseases, 10th edition (ICD-10)(21), and procedures using the Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures, 4th revision (OPCS)(22). Each patient is assigned a unique identifier (HESID), making it possible to study longitudinal patterns of care, such as rates of emergency readmission, as well as enabling the tracking of patients between hospitals.

Concerns about the accuracy and completeness of coding of diagnoses and procedures have been raised, but most NHS trusts in England submit good quality data to HES(23-28). To address these issues as far as possible, we have taken steps to ensure that the statistics we derive from HES are as valid and reliable as possible by:

1) carefully cleaning the data to remove duplicates and records not relating to an episode of care for a benign gynaecology condition
2) identifying units with missing data
3) making appropriate adjustments for case mix variation (see Chapter 4 & Appendix 1)

4. Methodology for deriving indicators

4.1. Cohort selection

The figures for emergency readmission, length of stay, and hysterectomy type are based on HES admitted patient data for the financial year 2015/16. Those for reoperation following MUS insertion (TVT/TOT) and pelvic organ prolapse (POP) repair are based on procedures conducted between 2011/12 and 2013/14 to allow a follow-up period of two years to adequately capture the reoperations forming the numerator. This is because, statistically, reoperation following these procedures is relatively rare, so using data from a three-year period allowed us to increase the sample size and therefore the statistical power to detect variation in rates of reoperation between trusts. In both cases, duplicate records were identified on the basis of HESID and date of admission. After removing duplicates, benign gynaecology records were defined as those which contained information about women and girls aged >11 years old with one or more specified procedure or diagnosis codes (code list available online). The youngest age among the patients extracted was 17 years. NHS trusts which performed at least 500 benign gynaecology procedures were included in the cohort. Trusts which performed fewer than 500 benign gynaecology procedures in the financial year 2015/16, or which closed during that financial year, were excluded, leaving 125 trusts. These ranged in size from 502 to 2625 benign gynaecology procedures per trust (mean:

1280 procedures, SD: 511). For each indicator, trusts conducting fewer than 5 procedures were excluded from the funnel plot.

4.2 Analysis and case mix adjustment

We present annual statistics for each English NHS trust that met our minimum data standards. Precise indicator definitions are provided in the corresponding results sections. Because clinical and demographic characteristics of patients can affect the demands placed on a service and the outcomes of care we take into account how similar patient groups are at each trust. Accounting for these factors, which are outside the control of care providers, is essential for fair and meaningful comparisons across trusts. We controlled for differences in case mix in several ways. First, the results of some indicators (length of stay and emergency readmission) are stratified by type of hysterectomy, as this may influence outcomes. For each indicator, demographic risk factors available in HES (age, ethnicity and area-level deprivation, Table 3) were included in a risk adjustment (multiple logistic regression) model. The type of hysterectomy indicators were also adjusted for diagnosis of prolapse, as this can influence the procedure performed (characteristics of women undergoing hysterectomy by condition are presented in Table 3). This regression model gave the probability of each woman having the outcome of interest on the basis of her characteristics. These probabilities were summed at the trust level to give the predicted rate. Risk adjusted rates were produced by dividing each trust's unadjusted rate by its predicted rate, and multiplying by the national mean rate. Further details in Appendix 1.

TABLE 2: DEMOGRAPHIC CHARACTERISTICS OF WOMEN IN THE HES BENIGN GYNAECOLOGY SAMPLE

	Categories	Frequency (%)
Age Group	15-24	3.1
	25-34	10.9
	35-44	23.5
	45-54	31.4
	55-64	14.2
	65-74	11.1
	75+	5.8
Ethnicity	White	78.3
	Asian	4.9
	Black African/Caribbean	4.2
	Other	5.3
	Unknown	7.3
Socioeconomic Deprivation (Index of Multiple Deprivation)	1 (least deprived)	18.6
	2	19.7
	3	20.4
	4	21.1
	5 (most deprived)	20.3

Relationships between indicators

We explored several potential relationships between indicators to see if these could inform indicator profiles at trust level. We conducted Pearson's Chi-squared tests to evaluate

whether trusts laying above or below the funnel limits for one indicator were more likely to lie outside the limits for another. We also calculated the number of trusts whose indicator values lay outside the funnel limits for more than one indicator. We explored the evidence for the following relationships:

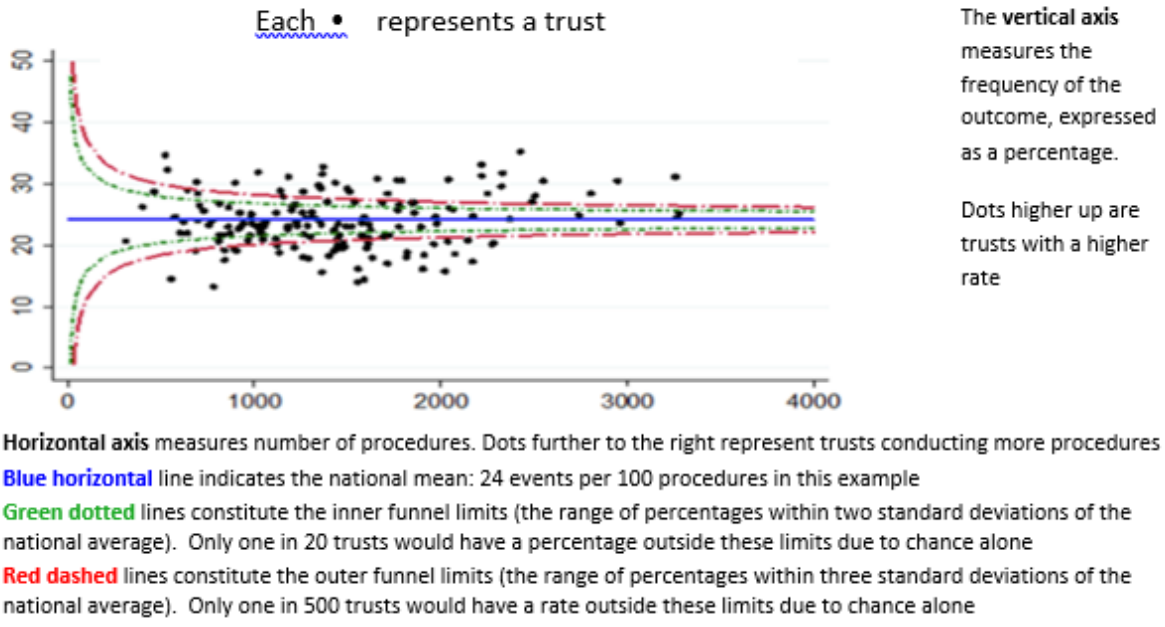
1. Are longer average lengths of stay for one hysterectomy type associated with longer stays for other types?
2. Are longer average lengths of stay associated with lower readmission rates?
3. Are shorter average lengths of stay associated with higher readmission rates?
4. Are higher rates of reoperation following MUS insertions associated with higher rates of reoperation following POP repairs?

We conducted analyses 2 and 3 for hysterectomy (all types combined, and by type), MUS insertions, POP repairs and endometrial ablation.

4.3 Presentation of data using funnel plots

We use funnel plots to highlight variation between trusts over and above what would be expected due to chance alone, taking their size into account(29). This is important because the amount a trust's indicator value may vary is influenced by random fluctuations, which are related to the number of procedures conducted (Figure 2).

FIGURE 2: HOW TO INTERPRET A FUNNEL PLOT



We have attempted to limit the impact of case-mix through risk-adjustment. However, this is feasibility work in terms of defining national indicators for benign gynaecology, so publishing results on a named-trust basis was not deemed appropriate and we do not use these plots to identify individual trusts as outliers with 'good' or 'poor' levels of performance.

A number of funnel plots show evidence of over-dispersion, which occurs when there is a greater level of variability than can be explained by chance or the existence of a few

outlying units(30). Explanations for over-dispersion can include variable data quality, limitations of the available risk adjustment methods and clinical uncertainty. We have not made statistical adjustments to widen funnel limits to deal with over-dispersion as our intention is to illustrate national variation rather than to identify outliers.

5. Results

The results for each of the indicators selected for this report are presented by procedure.

5.1. Hysterectomy Indicators

Of the 27,208 hysterectomies performed: 42.8% (n=11,634) were abdominal, 37.1% (n=10,091) vaginal and 20.2% (5,483) laparoscopic. In these indicators laparoscopic hysterectomy (LH) refers to total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy. We have combined LAVH with vaginal hysterectomy as LAVH incorporates surgical removal of the uterus through the vagina with most pedicles secured vaginally. In clinical practice we would expect readmission rates and lengths of stay for LAVH and VH to be similar. (The codes used to define procedures are provided in Appendix 1). The characteristics of women having hysterectomy are in Table 3.

TABLE 3: CHARACTERISTICS OF WOMEN UNDERGOING HYSTERECTOMY BY CONDITION

	Total	Prolapse (N81)	Endometriosis (N80)	Menstrual (N92-N94)	Fibroids (D25)
Total (n)	27,208	8,234	4,550	9,283	11,615
Average age					
Mean (SD)	52.0 (12.0)	61.7 (12.1)	49.2 (9.6)	44.6 (6.0)	50.5 (9.3)
Age group			**% (n)**		
15-24	-	-	-	-	-
25-34	3.8	1.4	3.1	6.3	0.8
35-44	24.1	7.7	27.4	38,6	22.4
45-54	39.7	19.7	48.9	52.7	55.4
55-64	14.0	25.4	11.4	2.0	11.1
65-74	12.4	30.9	6.8	0.3	7.2
75+	5.8	15.0	2.5	0.1	3.1
Hysterectomy type			**% (n)**		
Abdominal	42.8 (11,634)	3.7	45.1	51.7	60.6
Vaginal	37.1 (10,091)	93.3	29.5	19.2	20.8
Laparoscopic	20.2 (5,483)	3.0	25.5	29.1	18.6

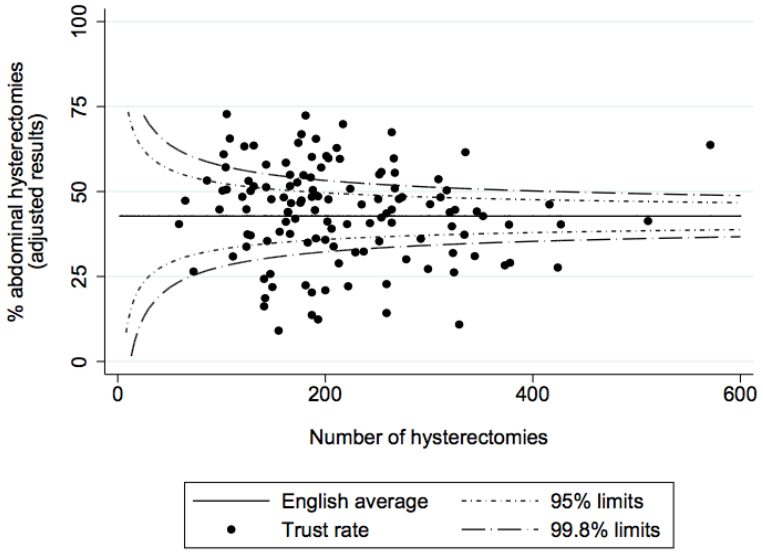
Footnotes: cells with ≤10 individuals suppressed. Percentages presented unless otherwise specified. Laparoscopic=total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy. Vaginal hysterectomy here includes laparoscopically assisted vaginal hysterectomy.

Construction of the hysterectomy type indicators

Definition	Numerator	Denominator
Proportion performed abdominally	Abdominal hysterectomies	All hysterectomies (any type)
Proportion performed vaginally	Laparoscopic hysterectomies	All hysterectomies (any type)
Proportion performed laparoscopically	Laparoscopic hysterectomies	All abdominal and laparoscopic hysterectomies

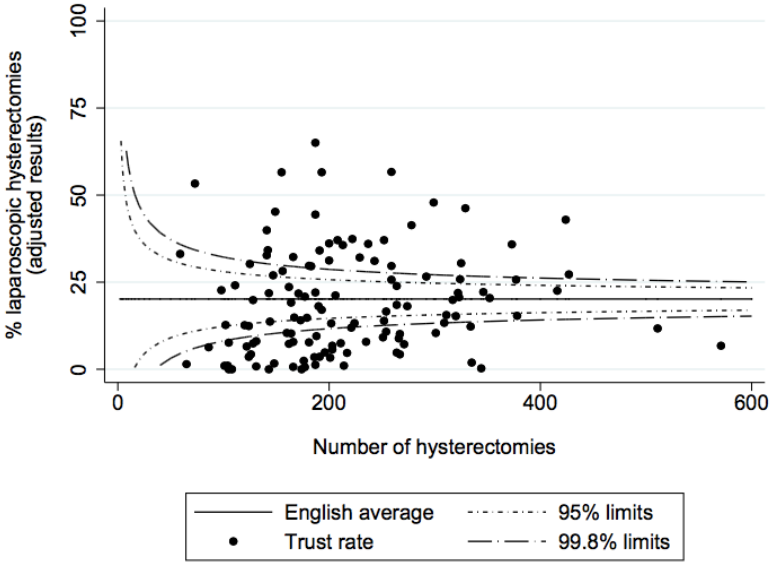
Footnotes: Laparoscopic=total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy. Vaginal hysterectomies here include laparoscopically assisted vaginal hysterectomies.

1. *Proportion of all hysterectomies performed abdominally*



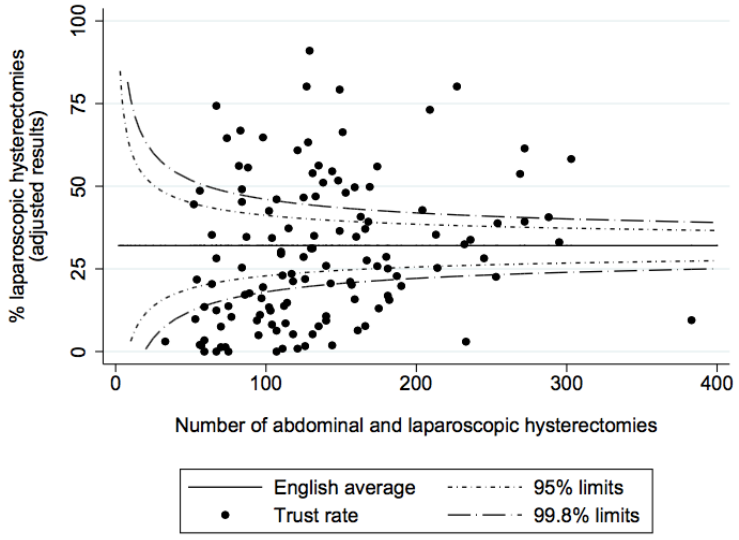
The mean proportion of hysterectomies performed abdominally was 42.8%. The mean among trusts in the lowest decile for this indicator was 17.4%, compared to 66.5% among trusts in the highest decile.

2. *Proportion of all hysterectomies performed via the laparoscopic route*



The mean proportion of hysterectomies performed that were laparoscopic was 20.2%. The mean among trusts in the lowest decile for this indicator was 0.6% vs. 49.7% among trusts in the highest decile

3. *Proportion of abdominal/laparoscopic hysterectomies performed via the laparoscopic route*



Amongst abdominal and laparoscopic hysterectomies, the mean proportion that were laparoscopic was 32.0%. The mean among trusts in the lowest decile for this indicator was 1.1% vs. 72.1% among trusts in the highest decile.

Length of stay following hysterectomy

Construction of the indicator

Definition	Numerator	Denominator
Proportion of women staying ≥1/≥2/≥3 night(s) after hysterectomy (H, AH, VH, LH)	Women staying ≥1/≥2/≥3 night(s) following a hysterectomy (H, AH, VH, LH)	H, AH, VH, LH (respectively)

H=Hysterectomy (all types combined); AH=Abdominal hysterectomy; VH=Vaginal Hysterectomy (including laparoscopically assisted vaginal hysterectomy); LH=Laparoscopic hysterectomy=total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy;

We present a number of different length of stay thresholds (≥1/≥2/≥3) to allow comparisons between the different types of hysterectomy.

Assessment of data quality

Procedures missing a date of surgery (n=4) were excluded as length of stay could not be calculated.

Results

Amongst all hysterectomies, the average proportion staying: ≥1 night was 74.0%, ≥2 nights was 35.5% and ≥3 nights was 15.3%. The means among the lowest and highest decile of trusts were: ≥1 night 40.2% vs. 97.4%), ≥2 nights 14.6% vs. 63.1% and ≥3 nights 5.4% vs. 29.0%. The average proportion staying: ≥1 night, ≥2 nights and ≥3 nights by type of hysterectomy is shown in Table 4.

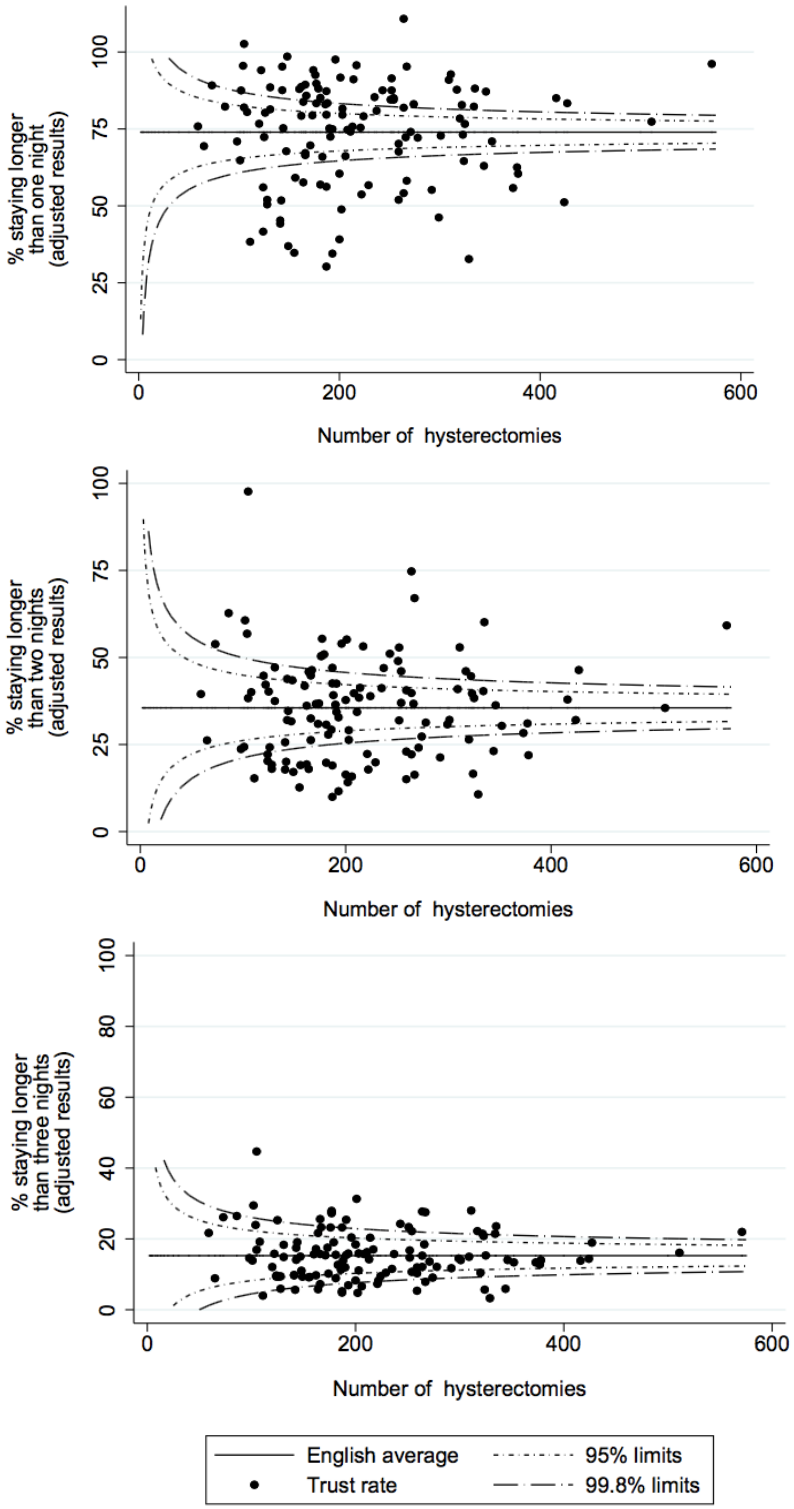
TABLE 4: LENGTH OF STAY BY TYPE OF HYSTERECTOMY

	Type of hysterectomy	National mean (%)	Mean in lowest decile (%)*	Mean in highest decile (%)*
Length of inpatient stay	≥3 nights			
	H[i]	15.3	5.4	29.0
	AH[ii]	24.2	8.7	39.4
	VH[iii]	7.2	1.0	21.6
	LH[iv]	4.6	0.0	18.4
	≥2 nights			
	H[i]	35.5	14.6	63.1
	AH[ii]	55.8	25.2	78.4
	VH[iii]	21.1	4.8	51.2
	LH[iv]	12.8	0.0	36.3
	≥1 night			
	H[i]	74.0	40.2	97.4
	AH[ii]	92.0	68.8	97.4
	VH[iii]	67.4	26.4	93.7
	LH[iv]	41.2	15.4	76.0

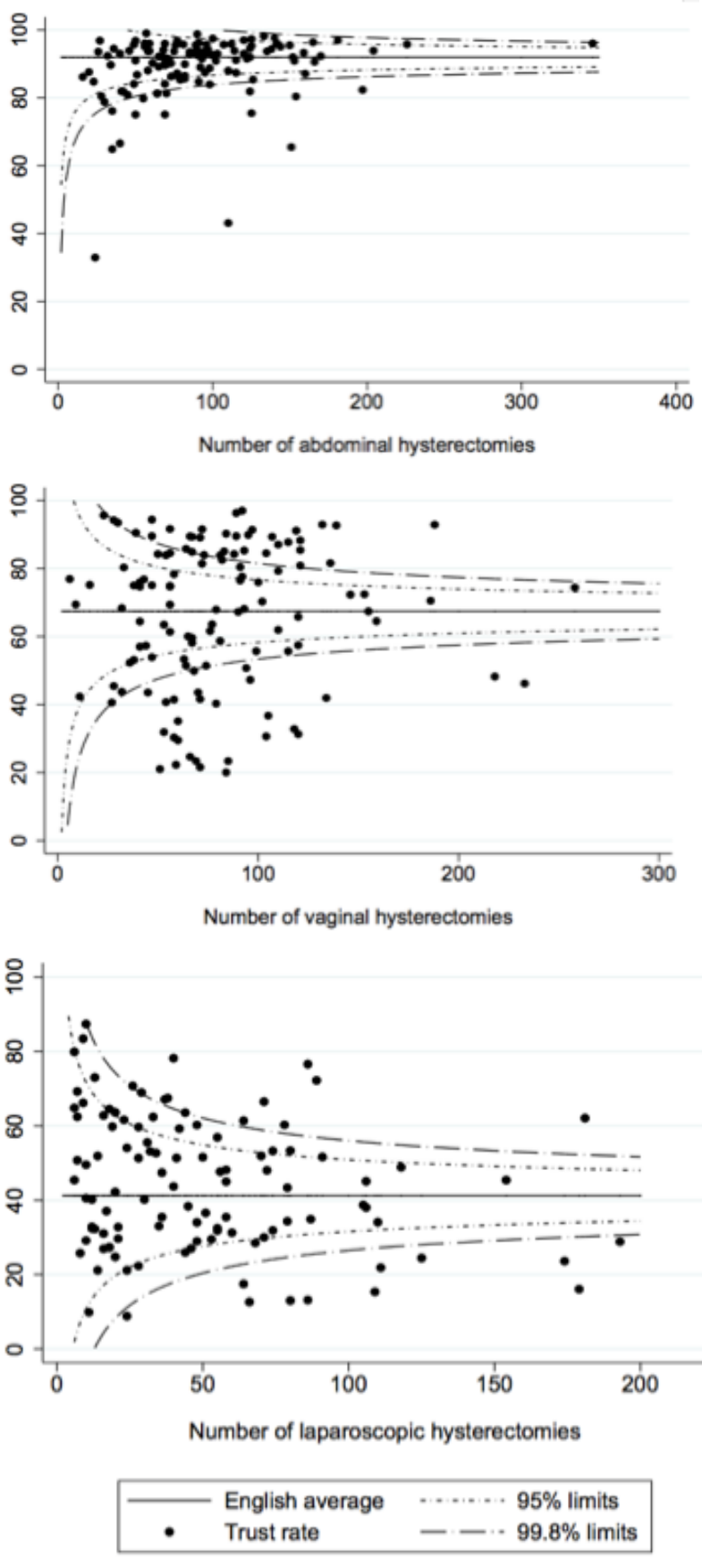
*Footnote: *After adjustment for demographic and clinical factors available in the dataset. [i]H=Hysterectomy (all types combined); [ii]AH=Abdominal hysterectomy; [iii]VH=Vaginal Hysterectomy (including laparoscopically assisted vaginal hysterectomy); [iv]LH=Laparoscopic hysterectomy=total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy*

We also investigated whether longer average lengths of stay for one hysterectomy type were associated with longer lengths of stay for other hysterectomy types, finding that trusts lying above the outer funnel limits for one hysterectomy route, were more likely to lay above the outer funnel limits for both of the other hysterectomy routes, for all three length of stay thresholds (≥1 night, ≥2 nights and ≥3 nights).

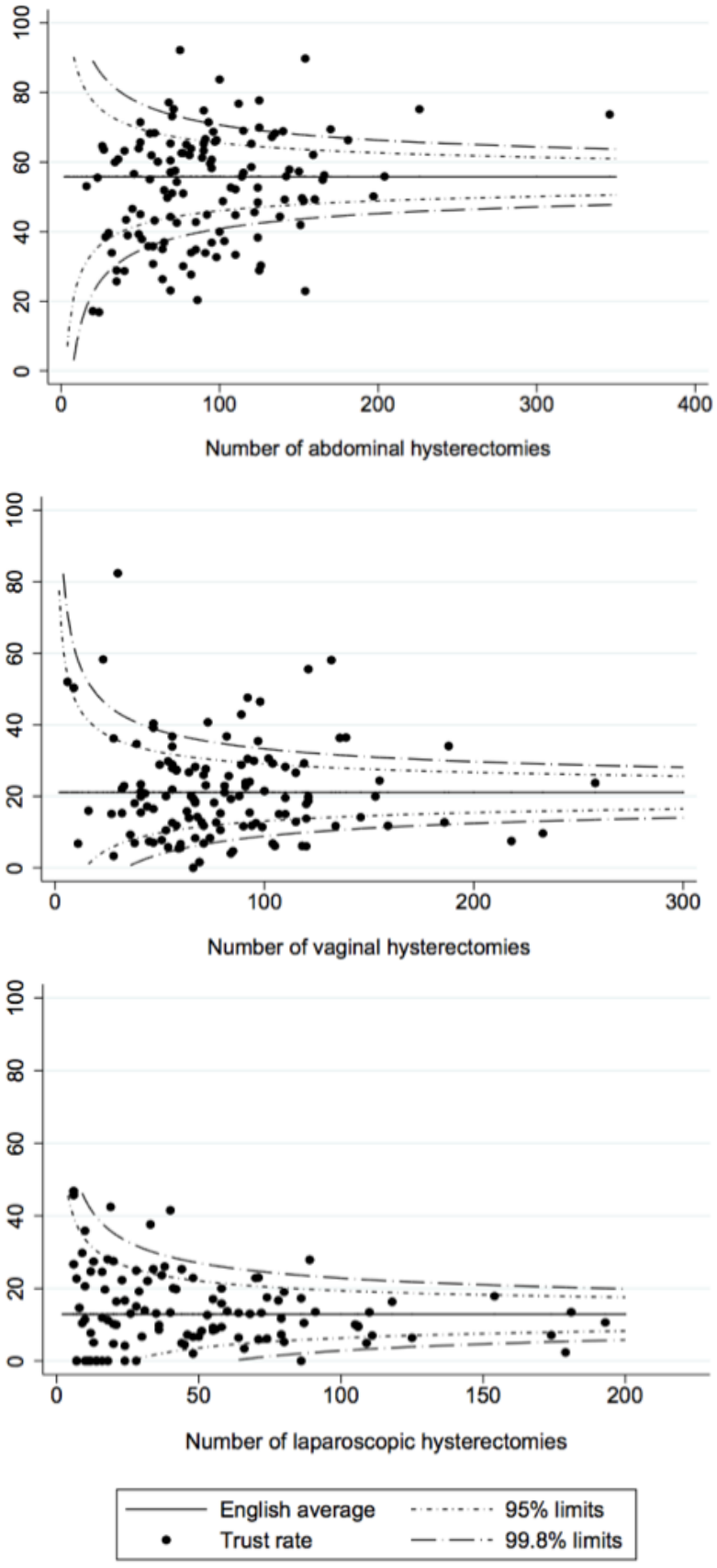
Length of stay following hysterectomy (all types combined)



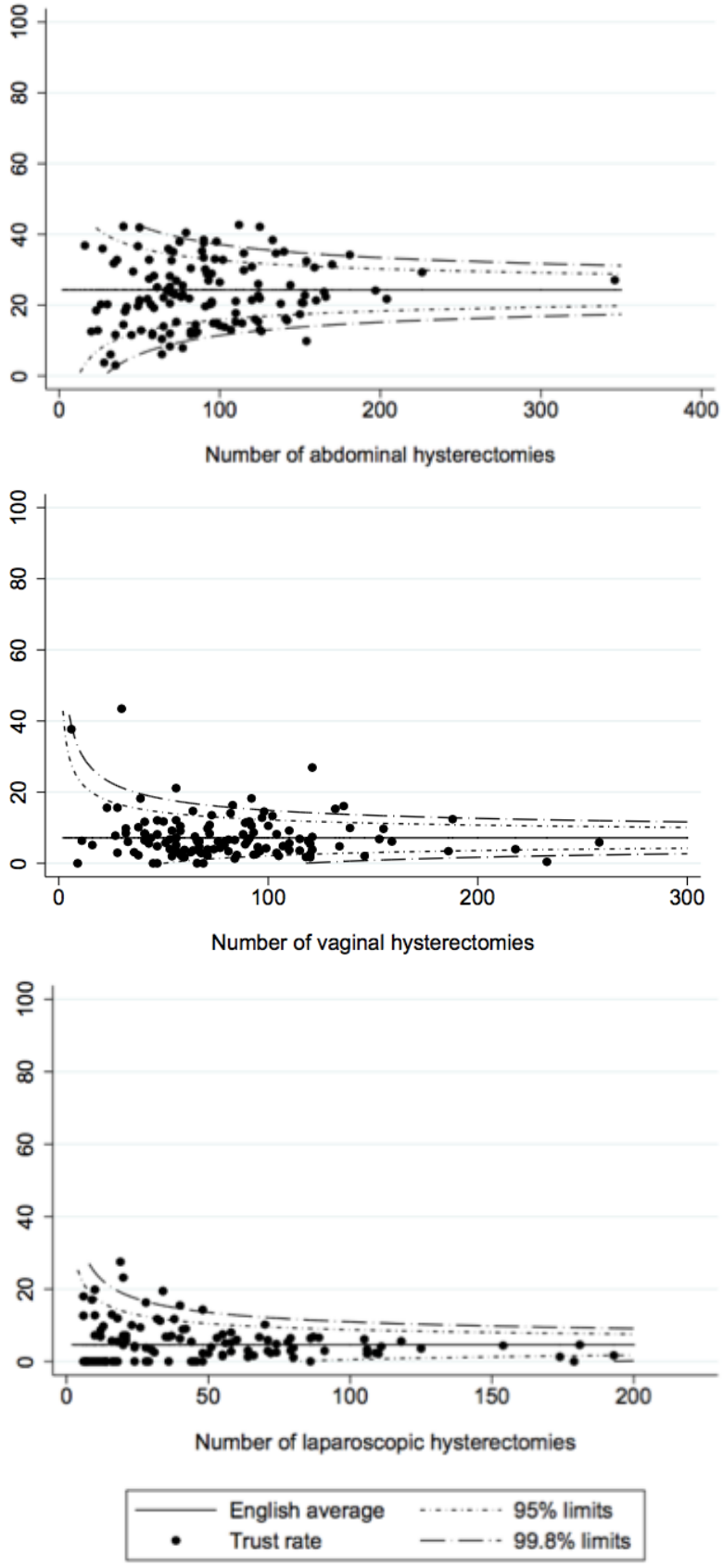
Length of stay more than one night following hysterectomy



Length of stay >2 nights following hysterectomy



Length of stay >3 nights following hysterectomy



Emergency readmission within 30 days of hysterectomy

Construction of the indicator

Definition	Numerator	Denominator
Proportion of women readmitted to hospital as an emergency within 30 days of hysterectomy (H, AH, VH, LH)	Emergency* readmission to any NHS hospital within 30 days of discharge after a hysterectomy (H, AH, VH, LH)	H, AH, VH, LH (respectively)

*H=Hysterectomy (all types combined); AH=Abdominal hysterectomy; VH=Vaginal Hysterectomy (including laparoscopically assisted vaginal hysterectomy); LH=Laparoscopic hysterectomy=total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy.** Emergency admissions were defined as any unplanned inpatient admission, referred via A&E, a GP, a consultant outpatient clinic or any other means. ***Note**: this indicator is restricted to emergency readmissions to secondary care. As such it does not capture non-emergency readmissions, or attendance at primary or secondary care within 30 days but without readmission, such as for vault haematoma treated with antibiotics.*

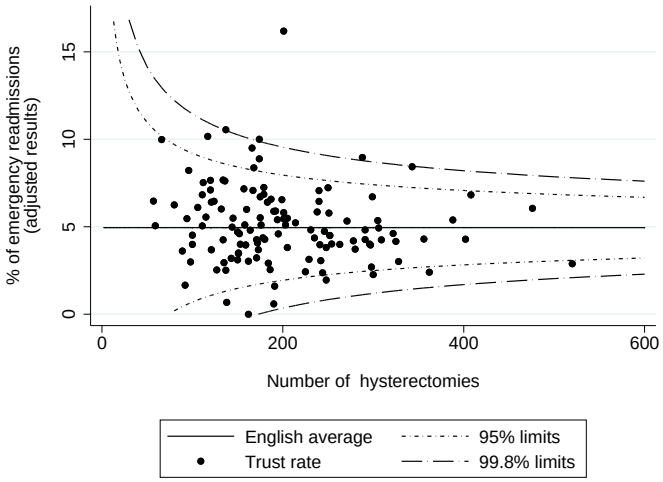
Assessment of data quality

Procedures missing a date of surgery (n=4) were excluded as readmission within 30 days could not be calculated.

Results

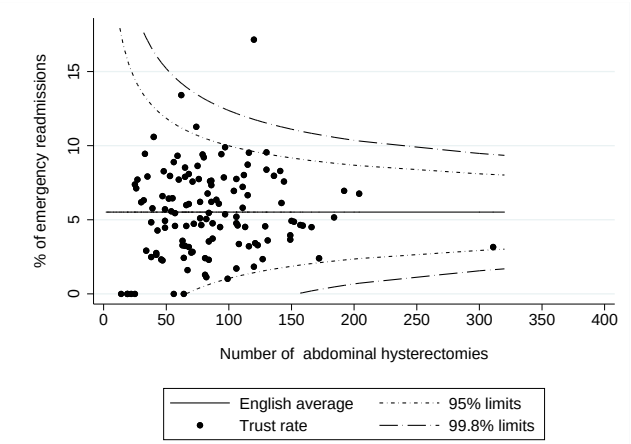
Emergency readmission within 30 days of hysterectomy (all types combined)

Amongst all hysterectomies, the mean rate of emergency readmission within 30 days was 5.0%. The mean rate was 1.8% among trusts in the lowest decile, and 9.7% among trusts in the highest decile, after adjustment for case-mix variation. The distribution of primary diagnoses for emergency readmission within 30 days is shown on the right below (The main reasons for emergency readmission were the same for each type (data not shown).

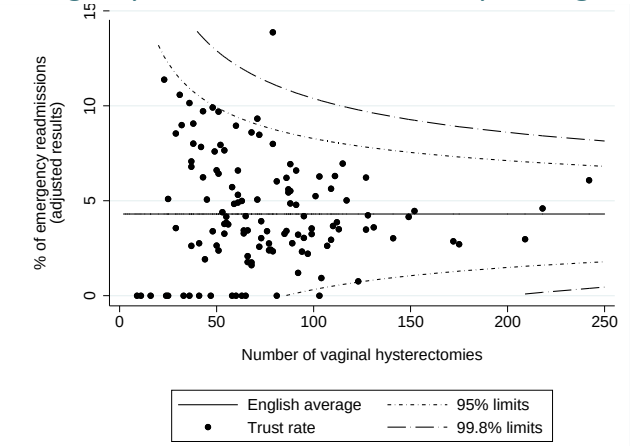


Primary diagnosis	Frequency among those readmitted as emergencies (%)
Haemorrhage	298 (21.2)
Infection	290 (20.6)
Constipation	73 (5.2)
Other and unspecified abdominal pain	71 (5.1)
Urinary tract infection	68 (4.8)
Pain localised to other parts of lower abdomen	62 (4.4)
Footnote: Only the most commonly recorded primary diagnoses associated with readmission are presented so the column does not sum to 100%	

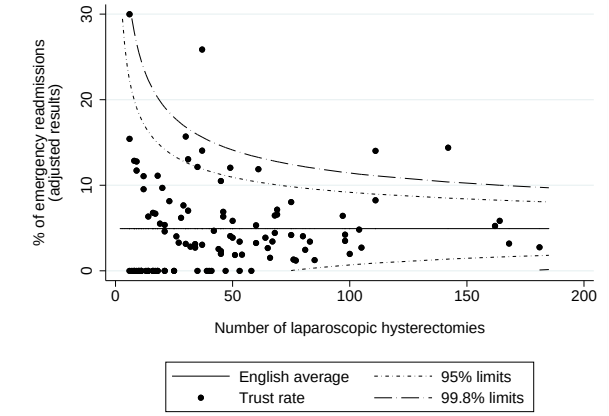
Emergency readmission within 30 days of abdominal hysterectomy



Emergency readmission within 30 days of vaginal hysterectomy



Emergency readmission within 30 days of laparoscopic hysterectomy



Amongst abdominal hysterectomies, the mean rate of emergency readmission within 30 days was 5.5%. The mean rate was 0.7% among trusts in the lowest decile, and 10.7% among trusts in the highest decile, after adjustment for case-mix variation. The mean rate of emergency readmission was lower for vaginal hysterectomies at 4.3% (10.0% among trusts in the highest decile). The mean rate of emergency readmission for laparoscopic hysterectomies was similar to that for abdominal hysterectomies at 4.9%, although the mean rate among trusts in the highest decile was 16.8%.

5.2. Mid-urethral Mesh Sling Indicators

Length of stay following MUS insertions (TVT/TOT)

Construction of the indicator

Definition	Numerator	Denominator
Proportion of inpatient spells ≥1 night after mid-urethral mesh sling (MUS) insertion	Women staying ≥1night after MUS insertion	All MUS insertions

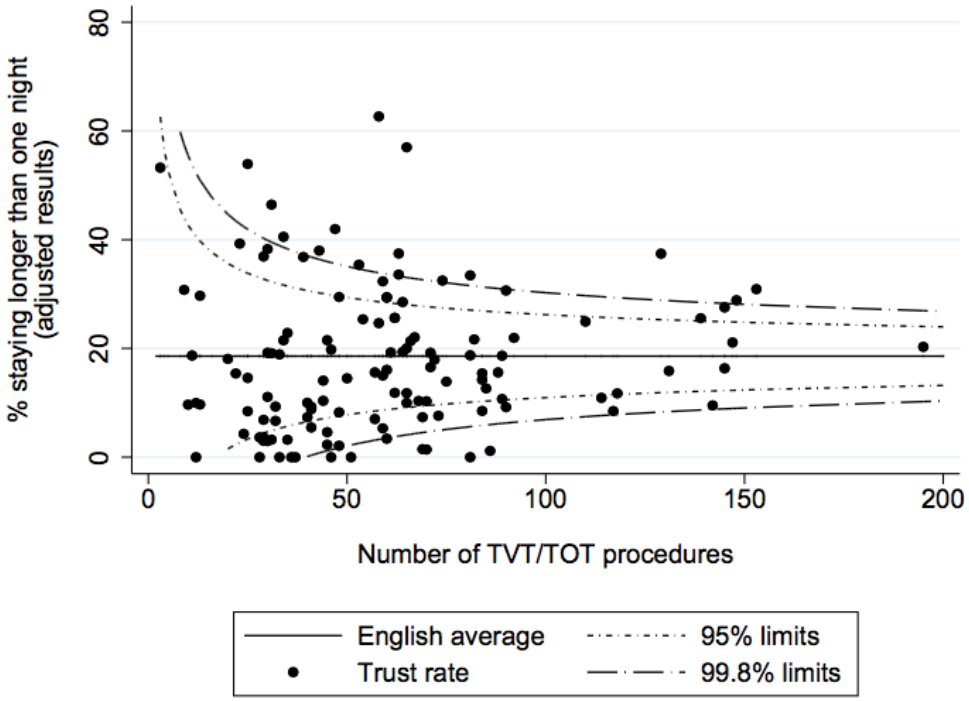
For procedures where women would not be expected to stay overnight (including MUS) we present the proportion staying ≥1 night.

Assessment of data quality

Procedures missing a date of surgery (n=11) were excluded as length of stay could not be calculated.

Results

Amongst MUS insertions (retropubic slings (TVT) and transobturator slings (TOT)), the average proportion staying ≥1 night following surgery was 18.6%. After adjustment for case mix variation, the average proportion staying ≥1 night among trusts with rates in the lowest decile for this indicator was 0.7%, compared to 45.5% among trusts in the highest decile.



Emergency readmission within 30 days of MUS insertion

Construction of the indicator

Definition	Numerator	Denominator
Proportion of women readmitted to hospital as an emergency within 30 days of MUS insertion	Emergency* readmission to any NHS hospital within 30 days of discharge after MUS insertion	All MUS insertions

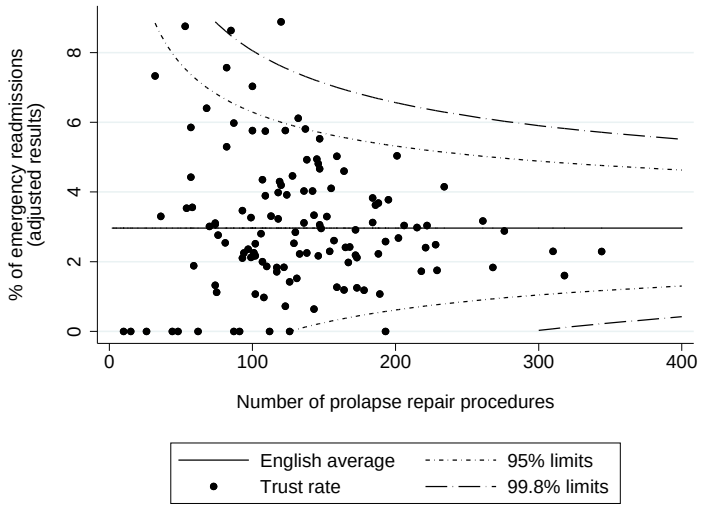
* Emergency admissions were defined as any unplanned inpatient admission, referred via A&E, a GP, a consultant outpatient clinic or any other means. ***Note**: as this indicator is restricted to emergency readmissions to secondary care it does not capture non-emergency readmissions, or attendance at primary or secondary care within 30 days but without readmission.*

Assessment of data quality

Procedures missing a date of surgery (n=11) were excluded as readmission within 30 days could not be calculated.

Results

Amongst MUS insertions (retropubic (TVT)/transobturator (TOT) route), the mean rate of emergency readmission to trust within 30 days was 2.4%. After adjustment for case mix variation the mean rate of emergency readmission in trusts in the highest decile was 8.7%. The primary reasons for emergency readmission following an MUS insertion are shown to the right, below.



Primary diagnosis	Frequency among those readmitted as emergencies (%)
Retention of urine	21 (15.1)
Infection	19 (13.7)
Urinary tract infection	16 (11.5)
Haemorrhage resulting from a procedure	15 (10.8)
Abdominal pain	7 (5)
Footnote: Only the most commonly recorded primary diagnoses associated with readmission are presented so the column does not sum to 100%	

Reoperation within 2 years of a mid-urethral mesh sling insertion

Construction of the indicator

Definition	Numerator	Denominator
Proportion of MUS insertions, repairs or removals within 2 years of previous MUS insertion	MUS insertions, repairs and removals within 2 years of previous MUS	All MUS insertions

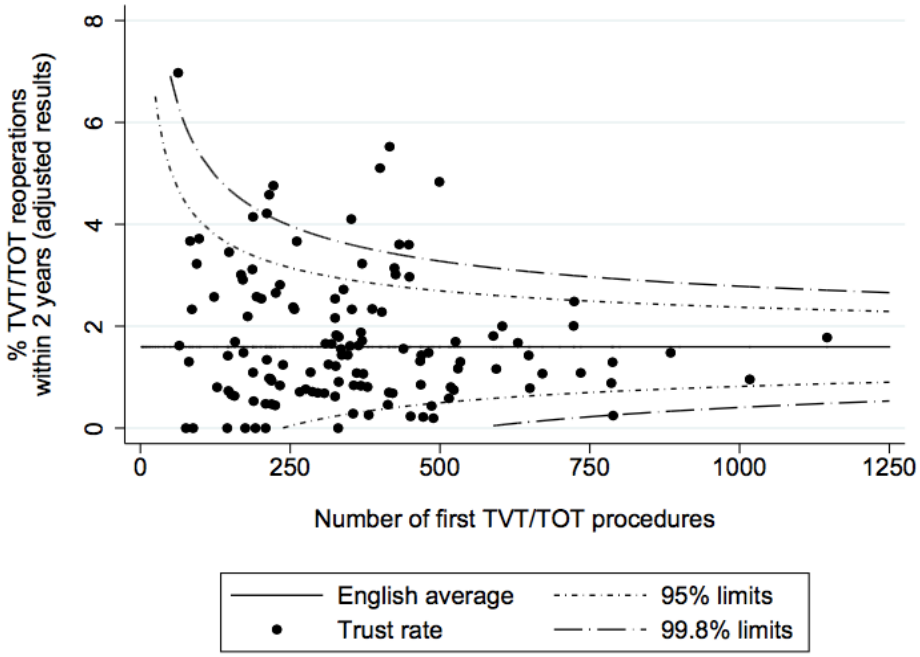
As described in the methods, the denominator for this indicator includes procedures conducted over a three-year period to increase the sample size and therefore statistical power to detect variation. The denominator therefore comprises ***all MUS insertions conducted between 2011/12 and 2013/14***, with a follow-up period of two years to capture reoperations forming the numerator.

Assessment of data quality

Twenty-two procedures were excluded as they were missing a date of surgery.

Results

Amongst MUS insertions (retropubic/transobturator route), the mean rate of reoperation for stress urinary incontinence within 2 years was 1.6%. After adjustment for case mix variation, the mean rate among trusts in the lowest decile of this indicator was 0.1%, compared to 4.6% among trusts with rates in the highest decile.



5.3. Prolapse Repair Indicators

Length of stay following pelvic organ prolapse (POP) repair

Construction of the indicator

Definition	Numerator	Denominator
Proportion of inpatient spells ≥2 nights after POP repair	Women staying ≥2nights after POP repair	All POP repair procedures

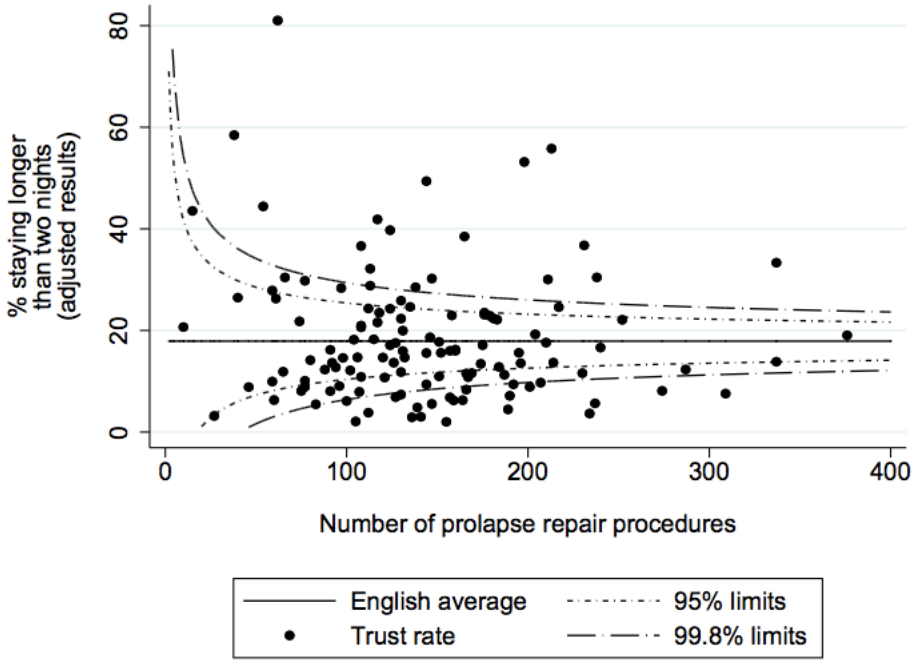
Women undergoing a POP repair would be expected to stay longer than those undergoing TVT or EA: we present the proportion staying ≥2 nights(31).

Assessment of data quality

Procedures missing a date of surgery (n=1) were excluded as length of stay could not be calculated.

Results

Amongst POP repair procedures, the average proportion staying ≥2 nights following surgery was 17.9%. After adjustment for case mix variation, the average proportion staying ≥2 nights among trusts with rates in the lowest decile was 4.1%, compared with 48.3% among trusts in the highest decile.



Emergency readmission within 30 days of POP repair

Construction of the indicator

Definition	Numerator	Denominator
Proportion of women readmitted to hospital as an emergency within 30 days of POP repair	Emergency* readmission to any NHS hospital within 30 days of discharge after POP repair	All POP repairs

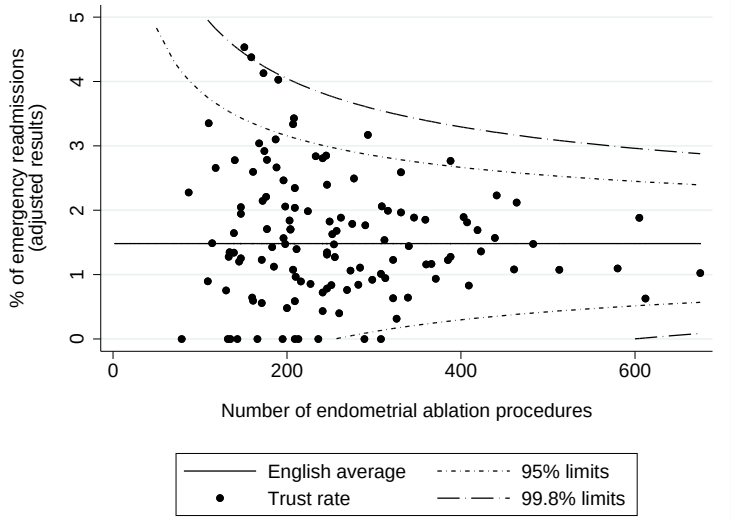
* Emergency admissions were defined as any unplanned inpatient admission, referred via A&E, a GP, a consultant outpatient clinic or any other means. ***Note**: as this indicator is restricted to emergency readmissions to secondary care it does not capture non-emergency readmissions, or attendance at primary or secondary care within 30 days but without readmission.*

Assessment of data quality

Procedures missing a date of surgery (n=1) were excluded as readmission within 30 days could not be calculated.

Results

Amongst POP repair procedures, the mean rate of emergency readmission to trust within 30 days was 3.0%. After adjustment for case mix variation, the mean rate among trusts in the lowest decile was 0.1% compared to 7.0% for trusts in the highest decile. The primary reasons for emergency readmissions following POP repair procedure are shown below.



Primary diagnosis	Frequency among those readmitted as emergencies (%)
Haemorrhage	94 (20.8)
Infection	68 (15)
Urinary tract infection	35 (7.7)
Retention of urine	29 (6.4)
Footnote: Only the most commonly recorded primary diagnoses associated with readmission are presented so the column does not sum to 100%	

Reoperation within 2 years of a POP repair procedure

Construction of the indicator

Definition	Numerator	Denominator
Proportion of reoperations (POP repair procedures and mesh removals or repairs) within 2 years of previous POP repairs	POP repair procedures and mesh removals or repairs (where mesh was used in the index POP procedure) within 2 years of POP repair	All POP repairs

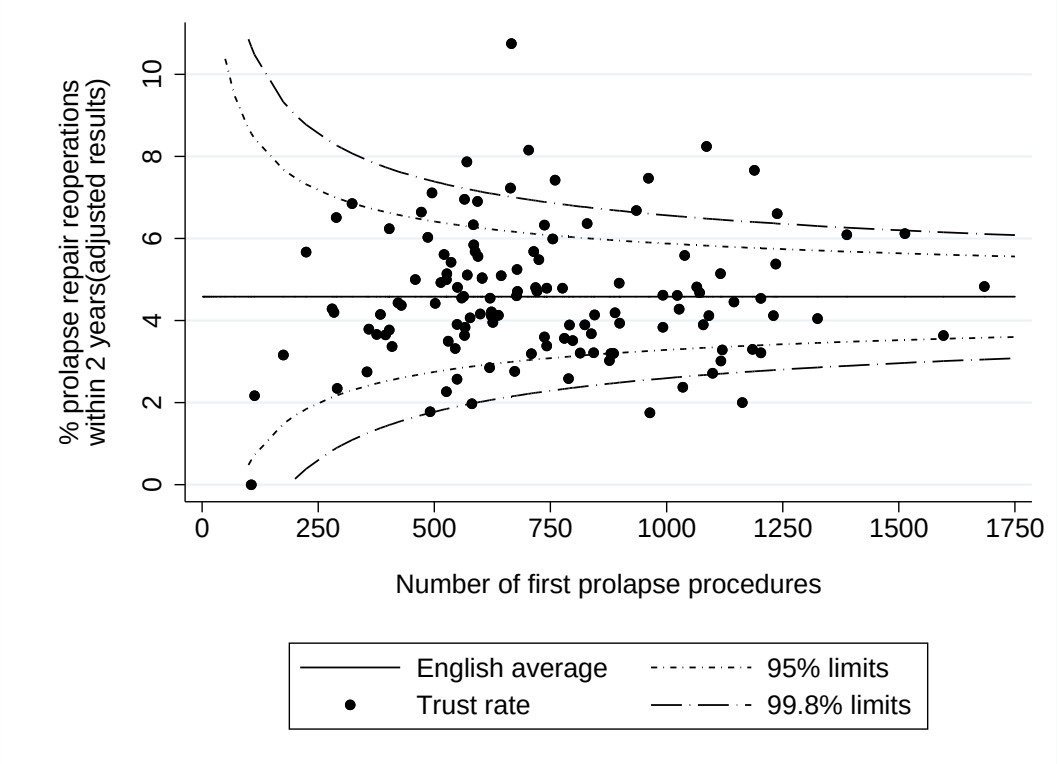
As described in the methods, the denominator for this indicator includes procedures conducted over a three-year period to increase the sample size and therefore statistical power to detect variation. The denominator therefore comprises ***all POP repairs conducted between 2011/12 and 2013/14***, with a follow-up period of two years to capture reoperations forming the numerator.

Assessment of data quality

Thirty-three procedures were excluded because they were missing a date of surgery (n=33).

Results

Amongst POP repair procedures, the mean rate of reoperation for prolapse within 2 years was 4.6%. After adjusting for case mix, the mean rate among trusts in the lowest decile of this indicator was 2.1%, compared to 7.7% among trusts with rates in the highest decile: more than a threefold difference.



5.4. Endometrial Ablation Indicators

Length of stay following endometrial ablation

Construction of the indicator

Definition	Numerator	Denominator
Proportion of inpatient spells ≥1 night after endometrial ablation (EA)	Women staying ≥1night after EA	All EA procedures

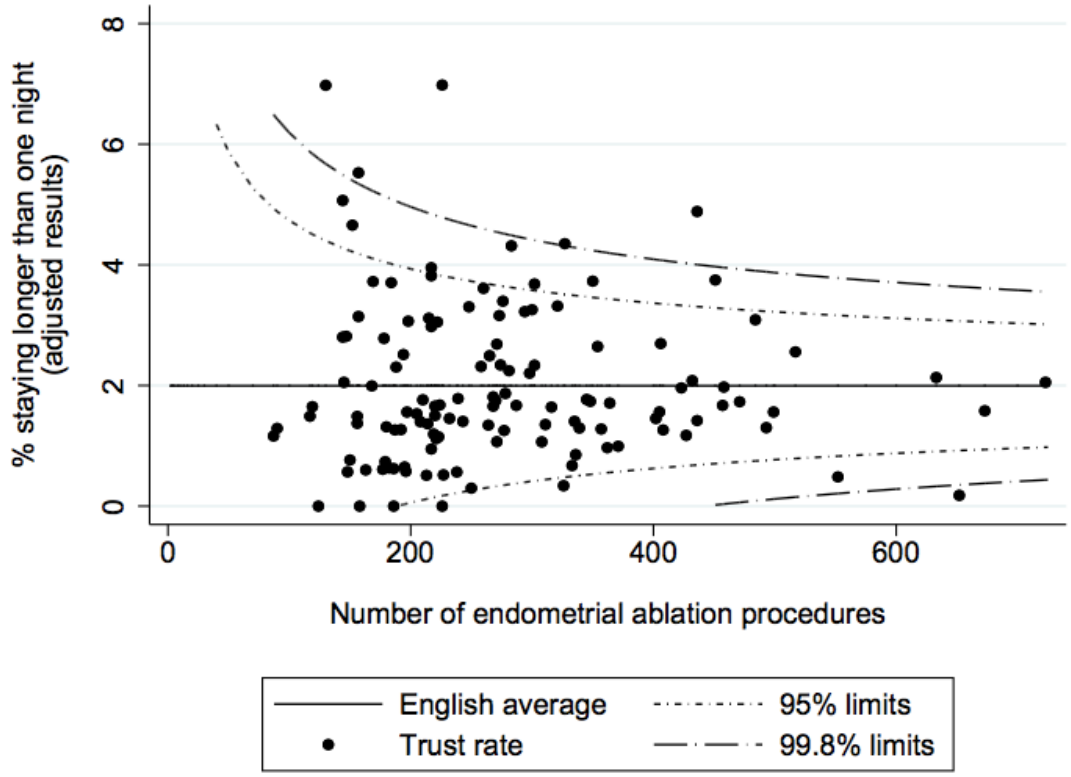
For procedures where women would not be expected to stay overnight (including EA) we present the proportion staying ≥1 night.

Assessment of data quality

Procedures missing a date of surgery (n=32) were excluded as length of stay could not be calculated.

Results

Amongst EA procedures, the average proportion staying ≥1 night following surgery was 2.0%. After adjustment for case mix, the average proportion staying ≥1 night among trusts with rates in the lowest decile for this indicator was 0.3%, compared to 4.8% among trusts in the highest decile.



Emergency readmission within 30 days of endometrial ablation

Construction of the indicator

Definition	Numerator	Denominator
Proportion of women readmitted to hospital as an emergency within 30 days of EA	Emergency* readmission to any NHS hospital within 30 days of discharge after EA	All EA procedures

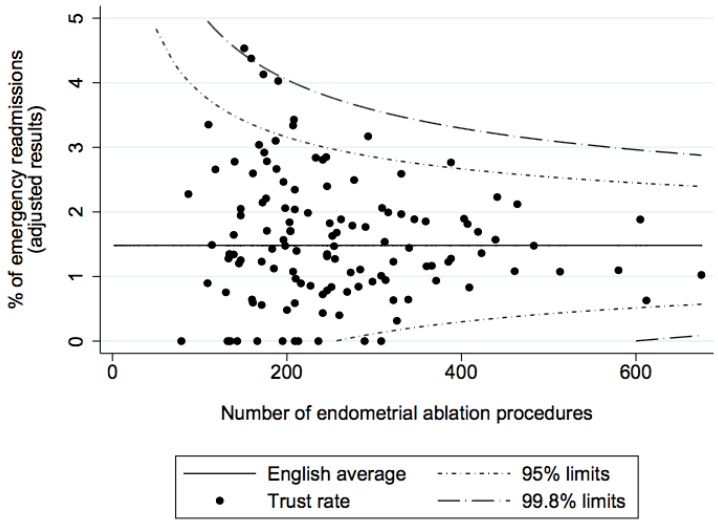
* Emergency admissions were defined as any unplanned inpatient admission, referred via A&E, a GP, a consultant outpatient clinic or any other means. ***Note**: as this indicator is restricted to emergency readmissions to secondary care it does not capture non-emergency readmissions, or attendance at primary or secondary care within 30 days but without readmission.*

Assessment of data quality

Procedures missing a date of surgery (n=32) were excluded as readmission within 30 days could not be calculated.

Results

Amongst endometrial ablation (EA) procedures, the mean rate of emergency readmission to trust within 30 days was 1.5%. After adjustment for case mix variation, the mean rate among trusts in the lowest decile was trust-level rates ranged between 0.1, compared with and 3.5% among trusts in the highest decile. The primary reasons for emergency readmissions recorded in HES are shown below.



Primary diagnosis	Frequency among those readmitted as emergencies (%)
Unspecified abdominal pain	39 (9.9)
Infection	38 (9.8)
Pain localised to other parts of the lower abdomen	35 (8.9)
Inflammatory disease of uterus	29 (7.3)
Footnote: Only the most commonly recorded primary diagnoses associated with readmissions are presented so the column does not sum to 100%	

5.5 Summary of results

TABLE 5: SUMMARY OF FINDINGS

Indicator	Subset of population	National mean (%)	Mean in lowest decile (%)*	Mean in highest decile(%)*
Proportion of all hysterectomies conducted abdominally	H[i]	42.8	17.4	66.5
Proportion of all hysterectomies conducted laparoscopically	H[i]	20.2	0.6	49.7
Proportion of abdominal/laparoscopic hysterectomies conducted laparoscopically	AH[ii]+ LH[iv]	32.0	1.1	72.1
Length of inpatient stay	≥3 nights			
	H[i]	15.3	5.4	29.0
	AH[ii]	24.2	8.7	39.4
	VH[iii]	7.2	1.0	21.6
	LH[iv]	4.6	0.0	18.4
	≥2 nights			
	H[i]	35.5	14.6	63.1
	AH[ii]	55.8	25.2	78.4
	VH[iii]	21.1	4.8	51.2
	LH[iv]	12.8	0.0	36.3
	≥1 night			
	H[i]	74.0	40.2	97.4
	AH[ii]	92.0	68.8	97.4
	VH[iii]	67.4	26.4	93.7
	LH[iv]	41.2	15.4	76.0
	MUS[v] (≥1 night)	18.6	0.7	45.5
	POP[vi] (≥2 nights)	17.9	4.1	48.3
	EA[vii] (≥1 night)	2.0	0.3	4.8
Emergency readmission within 30 days of surgery	H[i]	5.0	1.8	9.7
	AH[ii]	5.5	0.7	10.7
	VH[iii]	4.3	0.0	10.0
	LH[iv]	4.9	0.0	16.8
	MUS[v]	2.4	0.0	8.7
	POP[vi]	3.0	0.1	7.0
	EA[vii]	1.5	0.1	3.5
Reoperation within 2 years of MUS insertion	MUS[v]	1.6	0.1	4.6
Reoperation within 2 years of POP repair	POP[vi]	4.6	2.1	7.7

*Footnote: *After adjustment for demographic and clinical factors available in the dataset. For all indicators, women with malignancy were excluded. Additional exclusions were applied to each indicator on a case-by-case basis, as detailed in the main body of the report. [i]H=Hysterectomy (all types combined); [ii]AH=Abdominal hysterectomy; [iii]VH=Vaginal Hysterectomy (including laparoscopically assisted vaginal hysterectomy); [iv]LH=Laparoscopic hysterectomy=total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy; [v]MUS =Mid-urethral mesh sling for stress urinary incontinence; [vii]POP=pelvic organ prolapse repair; [vii]EA=Endometrial ablation.*

One in six trusts (n=21) lay outside the funnel limits for more than one of the indicators presented (not considering indicators further broken down by hysterectomy type).

6. Discussion

Hysterectomy type

Hysterectomy is one of the most commonly performed surgeries worldwide, and the most frequently performed major gynaecological surgical procedure(9), with the majority of hysterectomies conducted for a benign indication(10). Surgical approaches can be broadly categorised as abdominal (AH), vaginal (VH), and laparoscopic hysterectomy (LH)(9, 10).

There is controversy over variation in the proportion of hysterectomies in England carried out abdominally(32, 33). Data reported up to the year 2000 illustrated large variations in hysterectomy type between trusts, after accounting for the indication for hysterectomy (primary diagnosis) and age(33). Where technically feasible, laparotomy should be avoided and hysterectomy performed using a minimally invasive or vaginal route(34). There are a number of reasons for this. Vaginal hysterectomies are associated with lower mortality, post-operative morbidity and costs(10, 35). Focussing on evidence from randomised controlled trials (RCT), a Cochrane systematic review concluded that where technically feasible, vaginal hysterectomy was preferable to abdominal hysterectomy on the basis of faster recovery and fewer episodes of postoperative fever(10). The eVALuate study found that laparoscopic hysterectomy was associated with a higher risk of complications, and took longer to perform than abdominal hysterectomy(9). However, laparoscopic hysterectomy was also associated with a shorter recovery and better quality of life (at least in the short-term)(9).

In the 2015/16 financial year in English NHS trusts more than 40% of hysterectomies conducted for benign conditions were performed abdominally. Among abdominal and laparoscopic hysterectomies, most were abdominal with approximately only one in three done laparoscopically. This lower proportion of laparoscopic hysterectomies may reflect training backgrounds, or a narrowing in the indications for benign hysterectomy with increasing medicalisation of the treatment of menstrual disorders. The main purpose of laparoscopic hysterectomies should not be to supplant vaginal hysterectomies, but rather to replace abdominal hysterectomies, where clinically indicated, to reduce morbidity(36). A key indicator for the future will be to track whether rates of abdominal hysterectomies are declining because of a corresponding increase in the rate of laparoscopic hysterectomies.

The hysterectomy type indicators illustrated greater levels of variation than many other indicators in this report. Variation in the choice of hysterectomy type may also reflect variation in: clinicians' preferences; trust-level availability of surgeons to conduct laparoscopic hysterectomies; or trust policy, rather than quality of care. Trusts should review whether they provide the full range of hysterectomy route options, where clinically-indicated, so that decisions on hysterectomy route are not restricted by hospital factors.

Length of stay

Length of stay has been used to assess performance between providers for a wide variety of conditions and patient sub-populations(37, 38). Longer than expected inpatient stays following benign gynaecological surgery represent a deviation from the normal course of post-surgery recovery, with implications for both women and trusts. For procedures where

women would not normally be expected to stay overnight (MUS and EA) we present the proportion staying ≥1 night. Women undergoing a POP repair would be expected to stay longer than those undergoing TVT or EA: we present the proportion staying ≥2 nights(31). Three thresholds were included for hysterectomy (stays of: ≥1 night, ≥2 nights and ≥3 nights) to facilitate comparisons between the types of hysterectomy.

The lowest rates of longer than expected stays were observed following endometrial ablation, likely reflecting the minimally invasive, semi-automated and non-incisional nature of this procedure. Approximately one in six trusts had average lengths of stay beyond those recommended for MUS insertions (≥1 night) and POP repair (≥2 nights). Vaginal and laparoscopic hysterectomies were associated with shorter lengths of stay than abdominal hysterectomies, as expected. A higher proportion of trusts had longer inpatient stays following vaginal than laparoscopic hysterectomies (67.4% compared with 41.1% for ≥1 night and 21.1% compared with 12.8% for ≥2 nights). Length of stay following VH may be expected to vary depending on the indication for the procedure. We include vaginal hysterectomies both with and without concurrent POP repairs in the denominator for the VH indicators. Women undergoing VH for menstrual dysfunction are often discharged within 24 hours (as voiding is not an issue), whilst women undergoing a VH with a concurrent POP repair may be expected to stay longer for assessment of voiding. The higher proportion of trusts with longer inpatient stays following VH than LH might therefore be due to concurrent procedures related to the indication for the hysterectomy rather than the type of hysterectomy. This could be explored by trusts at a local-level. There was strong evidence for an association between length of stay for the different hysterectomy types. Trusts lying above the outer funnel limits for one type of hysterectomy were more likely to lay above the outer funnel limits for both the other types of hysterectomy, for all three length of stay thresholds (≥1, ≥2 and ≥3 nights). Variation in rates of longer stays may reflect variation in trust-level recovery and discharge procedures and policies, rather than quality of care (discussed further in the validity section below).

Emergency readmission

Rates of unplanned readmission within 30 days of discharge have also been used to assess performance between providers for a wide variety of conditions and patient sub-groups(37) as it represents an undesirable outcome. A majority of emergency readmissions related to the surgery will occur within this 30-day time frame. The mean rate of emergency readmission varied by procedure, with the highest observed following abdominal hysterectomy (5.5%) and the lowest following endometrial ablation (1.5%), reflecting the differing levels of invasiveness of these procedures. Substantial variation in the mean rate of emergency readmissions was seen for all procedures between trusts in the highest and lowest decile for these indicators. In addition, despite the wide variation in emergency readmission rates, very few trusts lay above the outer funnel limits for any of the procedures. This is likely to be related to statistical power; even in large datasets, where the number of events are relatively low, such as for emergency readmissions following benign gynaecology procedures, identification of trusts with higher or lower levels of performance is challenging(39). Trusts should locally audit their rates of, and reasons for, emergency readmission, to identify opportunities to improve care.

We also explored whether shorter lengths of stay were associated with higher readmission rates. We did not find evidence of this, but the number of trusts lying outside the funnel limits restricted our ability to detect an association.

Reoperation following MUS insertion or POP repair

Not all women with recurrent POP/UI following surgery will choose further surgery; many will choose conservative therapy or will live with their symptoms(16). However, reoperation is an undesirable outcome. We used a 2-year follow-up period because reoperation within this short timeframe can be considered to represent either technical failure, or issues with case or procedure selection. For prolapse, this may be related to not conducting a full prolapse assessment, or a multi-compartment repair.

Approximately one in every 20 women undergoing a POP repair procedure will undergo reoperation for prolapse within two years. A more than threefold difference in rates of reoperation for prolapse among trusts in the lowest and highest deciles suggests substantial geographic variation, however, less than 10% of trusts lay above the outer funnel limits, indicating that only a small proportion of this variation is above what would be expected by chance. Approximately one in every 60 women who received a MUS for SUI underwent reoperation within two years. We did not find evidence for an association between reoperation rates following MUS insertion and POP repair. However, the number of trusts lying outside the funnel limits was small, restricting our ability to detect an association.

At 4.6%, the national mean rate of reoperation following POP repair is lower than the 9-20% rates reported from recently published Scottish administrative hospital data(16). However, the latter rate was taken at 5-years of follow up, compared to the 2-year follow up here, and excluded POP repairs conducted at the same time as another procedure, and those in women who had undergone any POP procedure in the preceding 5 years. As such, the national mean rate of reoperation following POP repair procedures that we report of 4.6% may still be considered higher than desirable. Future work may consider a longer-time frame to capture a higher proportion of reoperations following POP repair procedures. However, this may also increase the proportion of reoperations that represent initial treatment of newly arising POP, rather than reoperation for technical failure or issues with case or procedure selection.

A pelvic organ prolapse can arise in the front wall of the vagina (referred to as the anterior compartment), the back wall of the vagina (known as the posterior compartment) or the uterus or top of the vagina (called the apical compartment), and can occur in more than one compartment at the same time. However, evaluation of the nature and extent of prolapse is subjective and surgeons may over or under-estimate the degree of prolapse; in some cases single compartment repairs may be conducted when multi-compartment repair might be more appropriate for symptom resolution. The rate of reoperation following POP repair at individual trusts may be amenable to reduction with an increase in the use of full prolapse assessments to ascertain the full extent and nature of the POP. This may lead to the need for multi-compartment repairs to be recognised at the first operation. Future work could compare the risk of reoperation among those who have a single compartment repair compared to those with a multi-compartment repair, and among those who undergo a POP repair that included a specific vault procedure, compared to those without.

Using HES data allows us to present variation in reoperation rates for prolapse after POP repair and for stress urinary incontinence after MUS insertion. However, reoperation rates are only part of the picture. They do not capture the many problems that women may have faced, such as pain, dyspareunia or recurrence of the original condition, and which may not have lead to further surgical treatment; some women who experienced these problems may have chosen conservative treatments, or no further treatment. Furthermore, these reoperation indicators do not distinguish between further continence procedures (a further MUS insertion or other procedures such as colposuspension, autologous slings of bulking agents) or further POP repairs respectively, and total or partial mesh removal procedures (following MUS insertions, or following POP repairs where mesh was used). Information on how often women had further surgery (a 'reoperation'), comprising either mesh removal or another continence or POP repair procedure, as defined in these indicators, will be of interest to women considering one of these procedures, and to clinicians. Further information on how many women who underwent an MUS insertion experience complications, require mesh removal and/or have a further continence procedure (and what type of procedure they receive) will be of interest to women considering continence surgery, and their surgeons, to help women to assess the balance of the potential risks and benefits of MUSs for SUI, and other continence procedures. To address this information gap, discussions are on-going about setting up a national prospective registry of mesh sling insertions for SUI and POP. Key considerations include how such a registry can provide information on the severity of the condition at the time of surgery (not available in administrative hospital data), and the issues that women face after treatment, which may initially lead to consultations in primary, rather than secondary, care.

In interpreting the results we also reflect on the four evaluation criteria used to select the indicators: validity, statistical power, technical specification and fairness (Figure 1).

Validity of performance indicators

The benign gynaecology indicators that can be derived from HES vary in terms of the sophistication of their construction, and the degree to which they are generic (emergency readmission within 30 days) or specific (the proportion of hysterectomies conducted abdominally). The challenge remains to establish the degree to which these indicator values represent valid measures of quality. For an indicator to be considered valid, a difference in the indicator should reflect a difference in the quality of care, with a specific direction reflecting better quality(18). In some situations, differences in the indicator may reflect quality of care, or may reflect other institutional factors such as variation in trust-level recovery and discharge procedures. For example, lower rates of longer stays or lower readmission rates could suggest that better care was received during the surgery episode, or for the latter, may indicate that women were better prepared for discharge by staff, for example by being given clear instructions about caring for surgical wounds to prevent infection, or good directions regarding medication regimes. A lower rate of longer stays may also point towards well organised enhanced recovery processes, whilst a low readmission rate may reflect well organised support services in the community once a woman is transferred home, or at other levels of the trust before a woman is readmitted as an inpatient. On the other hand, a low rate of longer stays could also be related to lower capacity, and the use of higher thresholds for keeping women as inpatients, or readmitting them, which may be the result of bed shortages. For POP repair indictors, trust policy on

post-operative catheter practices may influence length of stay and readmission rates. These conflicting potential interpretations, and the influence of variation in trust policy, therefore challenge the validity of length of stay and rate of emergency readmission as measures of quality, but they are widely used for benchmarking. The main difficulty in drawing conclusions based on this report, and others using administrative data to derive indicators, lies in defining which variation is unwarranted. Some level of variation is to be expected and indeed encouraged, representing patient-centred care. Eradication of all variation is not a reasonable aim(40). This report and the individualised results provided to trusts allow a 'comparative approach' to facilitate local investigations into and understanding of variation.

Technical specification and reliability of performance indicators

HES has a high case ascertainment and data completeness for many data items. For example, few episodes of care records had missing values for age and dates used to calculate indicators. For diagnoses and procedures, the level of missing data in HES is difficult to measure as the absence of a code for a diagnosis or procedure may reflect missing data, or that that diagnosis was not present, or the procedure was not conducted. However, there is evidence that primary diagnoses and procedures are well recorded, and that consistency between diagnoses and procedures is high(28, 41). In some cases, secondary diagnoses and procedures are less well recorded, which may impact on the extent to which further risk adjustment, beyond socio-demographic factors and primary indications for procedures included here, can be conducted(42).

Administrative hospital data such as HES are not collected for research or quality improvement. As such, not all data items required for specific indicators may be adequately captured. The technical specification of an indicator also needs to be sufficiently robust that it is not unduly influenced by records with poor or inconsistent data. In using administrative data, there is the possibility that indicators can be affected by omission or miscoding of diagnoses and procedures. However, a systematic review of coding accuracy in routine UK data found that 96% of primary diagnoses codes were accurate(41). There is also evidence that procedures codes are even more accurate than diagnosis codes, and that coding of admissions in HES is very complete, as is the distinction between emergency and non-emergency admissions(28, 41, 42). However, differences in coding practices may influence observed variation and trusts concerned about their emergency readmission rates should examine their admission method coding to explore if they have been recording non-emergency admissions as emergencies, as well as examining reasons for emergency readmission. The indicators used in this report mostly have robust technical specifications because they are calculated from OPCS-4 and ICD-10 codes.

Statistical power

It is necessary to consider the statistical power available to determine differences between trusts when evaluating a potential indicator. There are two factors that need to be taken into account: (1) the number of events that occur over a defined time period (the denominator) and (2) the frequency of the outcome (numerator). An advantage of using administrative data to develop indicators is their large sample size. For some procedures, such as all hysterectomy types combined, the size of the denominator can be large.

However, even in large datasets, where the number of events or procedures that form part of an indicator are low, identification of a trust with higher or lower levels of performance is challenging(39). Where the number of events (numerator) such as emergency readmission within 30 days of laparoscopic hysterectomy or endometrial ablation, is low, the impact of random fluctuations is increased. Using more than one year of data may be required to draw fair conclusions, though there would be a trade-off with the timeliness of the information. Low numbers can mask performance issues and lead to false complacency, which poses challenges for benchmarking trusts for quality improvement.

Power calculations allow a minimum number of procedures for individual trusts (or clinicians) to be set, below which a comparison against targets is not meaningful. Table 6 illustrates this point; we present hysterectomy indicators by type for clinical relevance, however, presenting indicators by type of hysterectomy introduces power issues due to smaller numbers of these procedures being conducted in some trusts. For example, the mean emergency readmission rate following vaginal hysterectomy was 4.3%. However, the median number of vaginal hysterectomies performed in 2015-16 per trust was 248, and no trusts performed enough vaginal hysterectomies annually to have 80% statistical power to detect a doubling in the rate of emergency readmissions. Care should therefore be taken to avoid over-interpreting indicators presented by hysterectomy route, or in other situations where the number of procedures performed is low, and the outcome is relatively rare, as statistical power will be limited. It is important to note that we excluded low 'volume' trusts performing fewer than 500 benign gynaecology procedures per year from the calculation of the indicators in this report. In terms of developing and analysing indicators a number of steps can be taken to increase statistical power. For example, indicators could be based on outcomes that are fairly frequent or derived from multiple years of data. For some procedures, such as hysterectomy, surgical routes could be analysed together, to increase the size of the denominator for each trust. However, this raises a different question: is it valid to consider these different surgical routes together?

TABLE 6: NUMBER OF PROCEDURES REQUIRED TO DETECT A RATE DOUBLE THE NATIONAL MEAN

Emergency readmission rate	**Median annual number of hysterectomies per trust**				**Number of procedures required to detect a rate double the national mean***		
	H	AH	VH	LH	60% power	70% power	80% power
	635	389	248	40			
3%					185	245	340
5%					95	132	179

**Power at 5% significance level. H=all types of hysterectomy, AH=abdominal hysterectomy, VH=vaginal hysterectomy (including laparoscopically assisted vaginal hysterectomy), LH=laparoscopic hysterectomy (total laparoscopic hysterectomy or subtotal laparoscopic hysterectomy)*

In this report we present results at the trust-level, however, in some instances, surgeons are required to report their individual-level results. As illustrated above for trust-level results, where the number of procedures a surgeon does per year is low, the chances of identifying those with less desirable rates of indicators is low, and so lack of evidence of poor performance should not be interpreted as evidence of acceptable performance(39). This may also pose challenges to individual surgeons in terms of revalidation.

Commissioners could consider whether there is an argument to combine gynaecology units

to increase the number of procedures performed at the trust-level, which could facilitate benchmarking of performance. This may also allow trusts to offer a comprehensive range of procedures, such as all types of hysterectomy, where they have appropriate referral pathways, and a team of surgeons with the full mix of skills. However, combining units may have the unintended consequence of increasing numbers of low 'volume' surgeons due to limited theatre time. Concern about low 'volume' of procedures is not just a statistical one; at the surgeon-level, performing low numbers of gynaecology procedures has been found to be associated with higher rates of intra- and post-operative complications(43). These considerations will need to be balanced locally.

Fairness

Not all variation in performance indicators will reflect variation in quality of care. Factors including random fluctuations, differences in data quality, and the case mix of patients between trusts may account for some of the variation observed(18). Patient characteristics may influence indications for procedures and treatments, as well as influencing outcomes. Indicators should therefore only be used for comparative purposes where adequate adjustment has been made for key case mix differences between populations of patients. Performance measurement should take into account the different populations and levels of disease severity treated by organisations through adequate risk adjustment. HES contains a number of variables commonly used in risk adjustment; age, ethnicity and socio-demographic deprivation are standard fields. In addition, co-existing diseases/conditions can be derived from HES diagnosis fields. A key strength of the indicator development process undertaken for this report is that adjustment for key case mix differences has been undertaken. This makes these indicators fairer than those used by some other initiatives(44). However, whilst we made efforts to make indicators as fair as possible by adjusting for key case mix differences, data were not available (in HES) on all factors that may influence the fairness of comparisons (for example, BMI and smoking status). Specifically, data were not available on size of the uterus or previous caesarean section or other abdominal surgery, both of which may influence preference for abdominal hysterectomy(33). Whilst it is possible to obtain information about obstetric history through linkage with historical HES data, this is only possible for deliveries since the start of data collection and the use of pseudonymised patient identifiers for longitudinal follow-up in HES (1st April 1997, with data completeness increasing over time, particularly since the 2004 introduction of “Payment By Results”(45, 46)). However, this means that information on obstetric history from historical HES data is not available for older women, a particular issue for many benign gynaecology conditions, where prevalence increases with age.

Expert Opinion

Many women come to see us with benign gynaecological complaints how content are we with the quality of the care we provide? How do our chosen processes of care and clinical outcomes compare with other hospitals across the country? I would argue that the answer to the latter question is that we simply do not know. This may be in part because in the field of obstetrics and gynaecology, scrutiny of clinical practice, including the organisation of services and the ensuing outcomes, has more often focussed on maternity care and gynaecological oncology. This report, examining the patterns of care in benign gynaecology, will go some way to addressing the current uncertainties and provide an impetus for self-reflection about our own practice and provision of services.

We are all aware of the deficiencies in the breadth and accuracy of routinely collected clinical data within our own hospitals and the wider NHS. However, without valid data we are in no position to know with any degree of certainty how we are doing and how we compare. This knowledge is essential if we are to make targeted, or in some instances wholesale, changes to enhance the care we provide.

The data from this report are derived from the Department of Health's Hospital Episodes Statistics (HES) data source and these routinely collected data, based on unique patient identifiers, do allow longitudinal observations to be made. The analysis of data have been adjusted for age, ethnicity and social deprivation, but the level of detail of routinely collected HES data precludes drilling down to a fine level of detail about care episodes and the conduct of a more sensitive analysis to explain unwarranted variations in practice and outcomes. The size of the national data set does however, allow a powerful and broad overview of current practices.

The data indicating care quality, included in this report are restricted to the practice and outcomes of hysterectomy, urogynaecology (specifically mid-urethral mesh tape insertion for urinary stress incontinence and pelvic organ prolapse repair) and endometrial ablation. In an ideal world, the report would be more comprehensive, evaluating more outcomes relating to the surgical interventions presented as well as interrogating other areas of benign gynaecological practice. However, the type of data collected and the low occurrence of specific indicators of interest, prevent a reliable assessment of relative differences across a broader range of gynaecological practice. However, the data contained within this report does provide a valuable snap shot of practice and outcomes for some of the most commonly performed procedures in contemporary gynaecology. These data can be used to 'indicate' where our work sits relative to others.

In 2015-6, 27,000 hysterectomies were performed in England predominantly for menstrual disorders and uterine fibroids. A laparotomic 'abdominal' approach to hysterectomy was the pre-eminent route selected. However, there was substantial variation across hospitals with rates of abdominal hysterectomy as low as 17% or as high as 66% of all hysterectomies performed. One third of all abdominal hysterectomies were performed laparoscopically but here the observed variation was even greater ranging from 1% to 72% and did not appear to be impacted upon by volume. This may reflect the relative lack of familiarity and proficiency with laparoscopic surgery and the paucity of training compared to conventional abdominal and vaginal hysterectomy. Vaginal hysterectomy was used in 37% of cases and not surprisingly was invariably the chosen route in the presence of prolapse. However, the

vaginal route was still utilised in 20-30% of hysterectomies for endometriosis, menstrual disorders and fibroids; broadly similar to the proportions for laparoscopic hysterectomy for these same indications. These rates may reflect the fact that 'laparoscopically assisted vaginal hysterectomy' was considered a vaginal hysterectomy rather than a laparoscopic one for the purposes of this report. Nevertheless, it is encouraging to see that less morbid, non-laparotomic routes were employed in approximately half of non-pelvic prolapse indications for hysterectomy. Advances in instrumentation, an emphasis on enhanced recovery and more acquaintance with laparoscopic hysterectomy should see this method become more prevalent in future reports. However, this shift in the preponderance of laparoscopic hysterectomy should be at the expense of abdominal hysterectomy rather than replacing less morbid vaginal hysterectomy.

Hysterectomy is emblematic of gynaecological surgery and we all have our own prejudices regarding how and when it should be done. More research is need to understand the relative risks and benefits of the route of hysterectomy within particular populations of women. However, the data in this report provide comparative information regarding length of stay (an indicator of post-operative recovery) and the need for emergency readmission (an indicator of surgical morbidity). It is encouraging to see that two thirds of women do not stay longer than two nights in hospital but this earlier average discharge is driven by vaginal and laparoscopic approaches, because more than half the abdominal hysterectomies stayed for two nights or more. If you are working in a hospital where your mean lengths of stay for a particular type of hysterectomy lie above the national upper decile, then an urgent look at your practices seems necessary. Indeed, trusts lying above the outer funnel limits for one type of hysterectomy (indicating that this result has only a 1 in 500 probability of being due to chance alone), were more likely to lay above the outer funnel limits for both the other types of hysterectomy. The increased length of stay may not reflect increased morbidity but rather practice routines and care pathways. With a drive towards 23 hour stay and enhanced recovery, these data are timely to inform such service development and audit its impact; for better or for worse. Readmission rates were broadly similar across hysterectomy subtypes with about one in five women readmitted within 30 days of the index procedure. Substantial variations in practice were again apparent, shining a light on the provision of care for this common procedure. If you are in a unit lying outside the funnel limits, such observations mandate the need for further examination of standard practices and the information acquired provides an opportunity to implement changes to reduce lengths of stay and readmissions.

Endometrial ablation is a semi-automated, relatively simple, common day-case or outpatient treatment for heavy menstrual bleeding, which is a recommended surgical treatment in the recently updated by NICE heavy menstrual bleeding guidelines (1). However, if it is so simple surely recovery times and post-operative morbidity should be equitable across NHS trusts? Not so, as the data reveal that the need to stay overnight varies from 0.3% to almost 5% of cases and readmission from 0.1% to 7.0%. These may seem relatively low rates, but for a common, quick and simple, minimally invasive procedure, the variation is surprising and again for those units with these outcomes substantially above the national means, an audit into current practice is necessary. Moreover, liaison with other, similar but better performing units in this area may help elucidate the underlying reasons for these differences.

The urogynaecological indicators reveal marked variation in length of stay following mid-urethral mesh tape insertion for stress incontinence and repair of pelvic organ prolapse. 10% of units admitted nearly half their women for at least one night post-procedure whereas the national average stay of one night or more was around one in five women. Why are there these differences? Do they reflect established practices, differing post-operative bladder protocols or surgical morbidity? Readmission following these urogynaecological procedures was reassuringly low (2-3%), but in the upper decile of units, readmission rates at 30 days were four to five times higher. It is not clear whether readmission correlates with earlier discharge but if you are a unit with relatively prolonged length of stays and readmission rates at the higher end of the national pattern then this should trigger a re-evaluation of your current service provision.

This report raises the profile of benign gynaecology and I believe will be warmly welcomed by gynaecologists practising in the UK. The development and analysis of indicators within a nationally collected data set provides us with a fantastic tool to benchmark the quality of our own local services, reflect on current practices and drive improvement in the gynaecological care we provide to our patients. It is revealing that one in six trusts lay outside the funnel limits (i.e. a one in 20 probability that the outlying result was due to chance alone) for more than one of the indicators evaluated. The relative impact of the changes we make to enhance clinical outcomes can be evaluated with the publication of subsequent reports on the patterns of benign gynaecological care. It will be interesting to see if and when, clinical practice and outcomes become more homogenous across England.

Reference

1. Heavy menstrual bleeding: assessment and management NICE guideline [NG88] (available at: http://www.nice.org.uk/guidance/ng88)

Professor T Justin Clark
Consultant Gynaecologist, Birmingham Women's and Children's Hospital, Birmingham, UK
Honorary Professor of Gynaecology, University of Birmingham, Birmingham, UK

7. Conclusions

Quality in healthcare is a multifaceted concept, not amenable to a single performance measure or simple metric. There is broad agreement that the key domains of quality are effectiveness, safety, capacity, patient-centeredness, equity, access and timeliness(47, 48). We undertook an extensive systematic review of the international literature to identify potential indicators to describe the practice and outcomes of benign gynaecology care. The development of a balanced set of indicators from those identified depends on many things, not least having ready access to complete and reliable data sources that contain the required information. Currently, the main data sources available are routinely collected administrative health datasets like HES. Many of the indicators identified in the literature could not be derived from HES so the indicators presented do not cover important aspects of care, such as service user experience and primary or outpatient care. As such, they should not be used to build conclusions regarding quality in the broadest sense of the term. However, we begin the process of defining those indicators that are methodologically robust, clinically valid and could be implemented nationally using these data. We demonstrate that, in the short-term, HES can provide detailed information on certain aspects of quality related to inpatient procedures for benign gynaecology conditions. This represents a key step towards the development of a balanced suite of indicators for monitoring the quality of benign gynaecology care. In England, the majority of benign gynaecology procedures take place in NHS hospitals and are therefore captured by HES, reducing the risk of selection bias.

In addition to the RCOG clinical indicators programme, a number of other national initiatives have been developed with the aim of improving the quality of care in gynaecology. These projects share important aims to provide healthcare providers and commissioners with meaningful, high-quality information that allows them to identify priority areas for improving outcomes for women, based on administrative hospital data. The RCOG clinical indicators programme aims to lead in this area, and support other initiatives to promote quality improvement, by systematically evaluating proposed indicators with a consensus group comprising clinicians and women with recent experience of English NHS benign gynaecology services, and including risk adjustment for case-mix. The fairness of indicators is particularly important and an area where this RCOG project adds to the work done by other quality improvement initiatives. Indicators should only be used for comparative purposes where key differences between populations of patients (case-mix) have been accounted for, and many other initiatives do not do this yet. To maximise the impact of indicators derived from administrative data for quality improvement, peer-to-peer discussion, in combination with the use of robust methodologies to adjust for population differences, may help to drive clinically-led discussion and quality improvement. There are calls for multi-disciplinary specialist groups which include Royal Colleges and patients, to keep performance measures under review(49).

There is a wide variation in values for some indicators between similar NHS organisations, in some cases, over and above what would be expected by random fluctuations. This variation both reflects and impacts upon the clinical uncertainty surrounding certain procedures, equity of access to services, health outcomes and the efficient use of NHS resources. Having access to comparative figures may act as a trigger to look at practice within an organisation or geographical area. This could in turn help to prevent potentially avoidable higher rates of

less desirable outcomes by learning from the experience of trusts with lower rates, leading to improved care. The way forward is discussed in more detail below.

The way forward

We have demonstrated that administrative hospital data can be used to develop indicators for benign gynaecology. In future we hope to include similar administrative hospital data from the Patient Episode Data for Wales (PEDW), the Information Services Division (ISD) in Scotland, and Hospital Inpatient Statistics (HIS) in Northern Ireland. There is an inherent lag in the availability of administrative data for service evaluation, audit and research purposes. Using the indicator definitions in this report with local data may provide timely information for clinicians and trusts. However, interpreting local data, with their smaller sample sizes, should be undertaken cautiously, given potentially limited statistical power to identify variation in adverse outcomes.

Ideally, a suite of indicators will not be overly restricted by limitations of the data source. The availability of more clinically detailed data will enable the improvement of existing indicators through refinement of risk adjustment models, as well as the development of new indicators to produce a more balanced picture of the quality of benign gynaecology care. Future work towards this could include linkage with other sources of clinical and user experience data where possible. Collecting information directly from women on the nature and severity of their condition(s) (before and after treatment), and any problems they face following treatment would facilitate the development of a more balanced indicator set that includes the patient perspective. In the meantime, at the trust-level, gynaecology service user groups may provide contextual information on patient-experience to help with the interpretation of local indicator results.

Through this robust indicator development process we aimed to increase the value that clinicians get from the data they routinely collect, and to contribute to the evidence that HES data can be meaningfully used for research and performance monitoring. We hope that this may lead to improvements in the use of the data routinely recorded in benign gynaecology. However, this is only a starting point in a complex process to reduce unwarranted variation. A programme of work is now needed to identify causes of variation at a local level, to begin to define acceptable levels of variation for each of the indicators and how unwarranted variation might be tackled. One in six trusts (n=21) lay outside the funnel limits for more than one of the indicators presented (not considering indicators further broken down by hysterectomy type). Further work is therefore also needed to understand relationships between indicators (the patterns of care within trusts), and to consider which indicators are best suited to support different aims (patient safety, performance assessment or quality improvement). We recognise that trusts are already actively engaged in efforts to reduce unwarranted variation, and this report is intended to act as an enabler to that process.

Appendices

Appendix 1. Additional detail on the data and analytical methodology

Data cleaning and indicator definitions

The basic unit recorded in HES is the finished consultant episode (the period of time a patient is under the care of one consultant). A 'spell' or admission is defined as the continuous period of time spent as a patient within one hospital from admission to discharge or transfer to another provider and may include >1 consultant episode.

For the purpose of this analysis, a benign gynaecology episode was defined as any record that contained valid information about a benign gynaecology procedure or diagnosis (full code list in online supplementary information). Duplicate records were identified on the basis of matching HESID and episode start date. The sample was restricted to women without a diagnosis of malignancy. Malignancy was defined using the ICD-10 codes for malignant neoplasms (C00-C75), in situ neoplasms (D00-D09), neoplasms of uncertain or unknown behaviour (D37-D48) and chemotherapy session for neoplasm (Z511), or the treatment specialty field in the HES database ('tretspef' 503 (gynaecological oncology)).

For the indicators, procedures were defined using the codes in Table A1.

Table A1: OPCS-4 codes used to identify procedures

Procedure	OPCS-4 codes
Hysterectomy	
Abdominal	Q07 (not with Y75.1 or Y75.2)
Vaginal (including laparoscopically assisted vaginal hysterectomy)	Q08
Laparoscopic (total or subtotal laparoscopic hysterectomy)	Q07 (with Y75.1 or Y75.2)
Mid-urethral mesh sling **insertions**	
Introduction of tension-free vaginal tape	M53.3
Introduction of transobturator tape	M53.6
Prolapse Repair Procedures	
Anterior colporrhaphy (non-mesh)	P23.2, P23.5 (not with Y02, Y36, Y37) (not with Q01 or Q08)
Anterior colporrhaphy (mesh)	P23.6 (not with Q01 or Q08)
Posterior colporrhaphy (non-mesh)	P23.3 or P23.4 (not with Y02, Y36, Y37) (not with Q01 or Q08)
Posterior colporrhaphy (mesh)	P23.7 (not with Q01 or Q08)
Sacrospinous fixation of vagina (non-mesh)	P24.4, P24.7
Vaginal vault repair (mesh)	P24.6
Open sacrocolpopexy (abdominal mesh)	P24.2, P24.5 (not with Y50.8 or Y75)
Vaginal hysterectomy for prolapse	Q08 (with ICD-10 diagnosis code N81)
Endometrial ablation	Q16, Q17

Emergency readmission was defined using the admission method field in HES ('admimeth' 21-24, 28, 2A, 2B and 2D). The discharge date from the readmission must be at least one day after the readmission date. Planned transfers are identified as follows: the admission date of the second spell is within plus/minus one day of the discharge date. Either the first spell has a discharge destination of 51 or 52, or the second spell has an admission source of 51 or 52, or the second spell has an admission method of 81. The readmission can be to any trust but is attributed to the trust where the initial procedure took place.

Reoperation following POP repair procedures and mid-urethral mesh sling insertions includes both mesh removals (where mesh was used in the index procedure), repairs, and further continence or prolapse repair procedures respectively. Reoperation following mid-urethral sling procedures was defined using the codes in Table A2. Reoperation following prolapse repair procedures was defined using the codes in Table A3. Reoperation was defined as a qualifying OPCS-4 code occurring during a hospital stay subsequent to, and with a discharge date within 2 years of, the index procedure date. Each of a patient's subsequent reoperations was counted separately.

Table A2: OPCS-4.7 codes used to identify reoperation after mid-urethral mesh sling insertions

Procedure description	OPSC-4.7 code
Further mid-urethral mesh sling (MUS) insertions	
Introduction of tension-free vaginal tape	M53.3 (excluding Y* codes listed as removal or repair)
Introduction of transobturator tape	M53.6 (excluding Y* codes listed as removal or repair)
MUS removal procedures	
Total removal of tension-free vaginal tape	M53.4
Partial removal of tension-free vaginal tape	M53.5
Removal of transobturator tape	M53.7
Unspecified tension-free vaginal tape removal	M53.3 AND * Y037: Removal of prosthesis from organ NOC * Y264: Removal of other repair material from organ NOC
Unspecified transobturator tape removal	M53.6 AND * Y037: Removal of prosthesis from organ NOC * Y264: Removal of other repair material from organ NOC
MUS repair procedures	M53.3 or M53.6 AND * Y031: Maintenance of prosthesis in organ NOC * Y033: Correction of displacement of prosthesis NOC * Y034: Other resitting of prosthesis in organ NOC * Y036: Adjustment to prosthesis in organ NOC * Y038: Other specified attention to prosthesis in organ NOC * Y039: Unspecified attention to prosthesis in organ NOC * Y265: Other attention to repair of organ NOC * Y302: Renewal of prosthesis in organ NOC * Y712: Secondary operations NOC * Y713: Revisional operations NOC * Y716: Second revisional operation NOC * Y717: Third or greater revisional operation NOC
Non-mesh SUI operations	
Bulking injections	
Endoscopic injection of inert substance into outlet of female bladder	M56.3
Other operations	
Abdominoperineal suspension of urethra	M51.1
Endoscopic suspension of neck of bladder	M51.2
Other specified combined abdominal and vaginal operations to support outlet of female bladder	M51.8
Unspecified combined abdominal and vaginal operations to support outlet of female bladder	M51.9

Table A3: OPCS-4.7 codes used to identify reoperation after prolapse repair procedures

Procedure description	OPSC-4.7 code
Further POP repair procedures	
Anterior colporrhaphy (non-mesh)	P23.2, P23.5 (not with Y02, Y36, Y37) (not with Q01 or Q08)
Anterior colporrhaphy (mesh)	P23.6 (not with Q01 or Q08)
Posterior colporrhaphy (non-mesh)	P23.3 or P23.4 (not with Y02, Y36, Y37) (not with Q01 or Q08)
Posterior colporrhaphy (mesh)	P23.7 (not with Q01 or Q08)
Sacrospinous fixation of vagina (non-mesh)	P24.4, P24.7
Vaginal vault repair (mesh)	P24.6
Open sacrocolpopexy (abdominal mesh)	P24.2, P24.5 (not with Y50.8 or Y75)
Vaginal hysterectomy for prolapse	Q08 (with ICD-10 diagnosis code N81)
Colpocleisis	P18
Anterior and posterior colporrhaphy and amputation of cervix uteri	P22
POP mesh removal/repair procedures	
Codes indicating lower genital tract paired with supplementary code indicating 'attention to prosthesis' or 'other repair of organ' to denote removal of prolapse mesh	P05 (with Y03 or Y26)
	P20 (with Y03 or Y26)
	P22.8 (with Y03 or Y26)
	P23.8 (with Y03 or Y26)
	P29 (with Y03 or Y26)
Footnote: PROSPECT validation data suggested that P22.8 and P23.8 codes paired with codes Y03 or Y26 are used to code prolapse mesh removal rather than (or possibly as well as) repeat colporrhaphy [16].	

Case mix adjustment

For each indicator, multiple logistic regression models were used to estimate the probability of a woman having had each intervention or outcome of interest on the basis of her age, ethnicity and level of socioeconomic deprivation. For the hysterectomy type indicators, prolapse diagnosis (ICD-10 code N81), was also included in the risk adjustment model.
Risk factor definitions:

- Age was defined using the age at start of episode (startage) field in HES and grouped into 7 categories: 1) 15-24, 2) 25-34, 3) 35-44, 4) 45-54, 5) 55-64, 6) 65-74 and ≥75.
- Ethnicity was defined using the ethnic category (ethnos) field in HES re-coded into 5 categories: 1) White, 2) Asian, 3) Black African/Caribbean, 4) Other and 5) Unknown.
- Deprivation was defined using a five-category indicator derived from the English Indices of Deprivation 2009 ranking of the English super output areas (areas of Census geography for presenting local statistical information). The categories were defined by partitioning the ranks of the 32,480 areas into quintiles and were labelled 1 (least deprived) to 5 (most deprived).

The probabilities of the intervention or outcome of interest for women who underwent their procedure at the same trust were then summed to give the trust's predicted rate. Risk adjusted rates for each trust were produced by dividing the trust's unadjusted rate by its predicted rate, and multiplying this ratio by the national mean.

The criteria used to evaluate models for measuring trust performance differ from those intended for the prediction of patient outcomes. For patient prediction, we want a system with the highest predictive ability. For hospital trust quality, we are seeking to measure a latent variable of quality and expect that differences in quality, which are unobserved, may account, at least in part, for the unexplained variation between hospital trusts that remains after adjusting for patient risk factors at presentation.

Appendix 2. Sources of indicators identified by the review

This table illustrates the sources of indicators that were identified in the literature review conducted in 2015, and where the indicators were also evaluated as potentially derivable from admitted patient HES data. Many other articles were identified by the review which proposed or reported on the use of a wide range of indicators throughout the benign gynaecology care pathway. The limited number of articles listed here illustrates that it was not possible to derive many of the indicators found in the literature from HES. In future, data linkage may facilitate the creation of a broader and more balanced suite of indicators, including measures of quality of life and experiences of care.

UK
Clinical performance indicators at hysterectomy. Attilakos et al. 2002. J. Obstet Gynaecol. Vol 22. Issue 1.
Introducing an acute gynaecology dashboard as a new clinical governance tool. Guha et al. 2013. Clinical Governance: An International Journal. Vol. 18. Issue 3.
Audit of quality indices in ectopic pregnancy. Chandreskar et al. 2014. RCOG Congress e-Poster
A novel early pregnancy assessment unit/Gynaecology assessment unit dashboard. An experience from a UK district general hospital. Wahba et al. 2015. Journal of Obstetrics and Gynaecology. Vol. 35
Standards if care provided by Early Pregnancy Assessment Units (EPAU): A UK wide survey. Poddar et al. 2011. Journal of Obstetrics and Gynaecology. Vol. 31
The NHS Outcomes Framework
Good practice in continence services. Department of Health. 2000.
Hospital Episode Statistics as a source of information on safety and quality in gynaecology to support revalidation. Royal College of Obstetricians and Gynaecologists. 2012.
Revalidation in Obstetrics and Gynaecology: Criteria, standards and evidence. Royal College of Obstetricians and Gynaecologists. 2002.
Measuring performance in hospital care: Length of stay in gynaecology. Leyland et al. 1997. Eur. J. Pub. Health. Vol 7. No. 2
Quality measures for the emergency obstetrics and gynaecology services. Drife. 2001. J R Soc Med. Vol 94. Suppl. 39.
Sweden
Medication Abortion as a quality indicator for regional comparisons in Sweden. Ljung et al. 2009. Am. J. Public Health. Vol 99, Issue 2.
United States of America
Quality assurance indicators and short-term outcome of hysterectomy. Gambone et al. 1997. International Journal of Obstetrics and Gynaecology. Vol. 35. Issue 4.
Outcomes in abdominal hysterectomy patients with benign disease. Use of physician-developed clinical protocols. D'Amato et al. 1998. J Reprod Med.
Trends in the national distribution of laparoscopic hysterectomies from 2003-2010. Lee et al. 2014. KJ Minim Invasive Gynecol. Vol 21. Issue 4
Defining success after surgery for pelvic organ prolapse. Barber et al. 2009. Obstet. Gynecol. Vol. 114. Issue 3.
Canada
Performance measurement in women's health: The Women's Health Report, Hospital Report 2001 Series, a Canadian experience. Magistretti et al. 2002. Women's Health Issues. Vol 12. Issue 6.
Reviews
Surgical Site Infections and the Surgical Care Improvement Project (SCIP): Evolution of National Quality Measures. Fry et al. 2008. Surgical Infections. Vol. 9. Issue 6.
Audit standards in ectopic pregnancy. Bajekal et al. 2000. J. Obstet. Gynaecol. Vol 20. Issue 2.
Quality Improvement Guidelines for Uterine Artery Embolization for Symptomatic Leiomyomas. (2010 and 2014 versions). Journal of Vascular and Interventional Radiology
Epidemiology and outcome assessment of pelvic organ prolapse. Barber et al. 2013. Int. Urogynecol. J. Vol. 24.
Footnote: Indicators in table 2 of the main report may have appeared in multiple sources.

Appendix 3. Indicator Consensus Group Membership

The consensus group that guided the indicator development process, and provided feedback on the report, consisted of gynaecologists and gynaecology nurses active in the English NHS, health services researchers and women with recent experience of English NHS benign gynaecology care (from the RCOG's Women's Voices Panel).

Chair: Jan van der Meulen
Professor of Health Services Research and Policy, London School of Hygiene and Tropical Medicine

Michelle Canton-Richards
RCOG Women's Voices Representative

Justin Clark
Consultant Gynaecologist and Honorary Professor of Gynaecology, Birmingham Women's and Children's NHS Foundation Trust

David Cromwell
Professor of Health Services Research and Policy, London School of Hygiene and Tropical Medicine; Director, Clinical Effectiveness Unit, Royal College of Surgeons

Dianne Crowe
Advanced Nurse Practitioner/Hysteroscopist, Northumbria Healthcare NHS Foundation Trust

Diane Danzebrink
RCOG Women's Voices Representative

Lucia Dolan
Consultant Gynaecologist Subspecialist in Urogynaecology, Belfast Health and Social Care Trust

Anita Dougall
Director, Clinical Quality, Royal College of Obstetricians and Gynaecologists and Qualified Midwife

Alison Elderfield
Head of the Lindsay Stewart Centre for Audit & Clinical Informatics, Royal College of Obstetricians and Gynaecologists

Ipek Gurol-Urganci
Assistant Professor, Health Services Research and Policy, London School of Hygiene and Tropical Medicine

Morag Hayes
RCOG Women's Voices Representative

Tim Hillard
Consultant Obstetrician and Gynaecologist, Poole Hospital NHS Foundation Trust

Debra Holloway
Gynaecology Nurse Consultant, Guy's & St Thomas' NHS Foundation Trust;
Chair of the Royal College of Nurses' Women's Health Forum

Jen Jardine
National Maternity and Perinatal Audit Clinical Fellow (Obstetrics), Royal College of Obstetricians and Gynaecologists

Hannah Knight
National Maternity and Perinatal Audit Lead, Royal College of Obstetricians and Gynaecologists

Edward Morris
Consultant Obstetrician and Gynaecologist, Norfolk and Norwich University Hospital
Vice President for Clinical Quality, Royal College of Obstetricians and Gynaecologists

Robert Sherwin
Consultant Obstetrician and Gynaecologist, Whittington Health NHS Trust
Clinical Lead, Getting it Right First Time: Obstetrics and Gynaecology

Helen Stevenson
Clinical Teaching Fellow, Birmingham Women's and Children's NHS Foundation Trust

Valerie Worth
RCOG Women's Voices Representative

References

1. Royal College of Obstetricians and Gynaecologists, London School of Hygiene & Tropical Medicine, Ipsos MORI. National Heavy Menstrual Bleeding Audit: First Annual Report. RCOG Press; 2011.
2. Royal College of Obstetricians and Gynaecologists, London School of Hygiene & Tropical Medicine, Ipsos MORI. National Heavy Menstrual Bleeding Audit: Third Annual Report. RCOG Press; 2013.
3. Cromwell D, Mahmood T, Templeton A, van der Meulen J. Surgery for menorrhagia within English regions: variation in rates of endometrial ablation and hysterectomy. BJOG : an international journal of obstetrics and gynaecology. 2009;116(10):1373-9.
4. Kiran A, Geary R, Gurol-Urganci I, Cromwell D, Bansi-Matharu L, Shakespeare J, et al. Sociodemographic differences in symptom severity and duration amongst women referred to secondary care for menorrhagia in England and Wales: a cohort study from the National Heavy Menstrual Bleeding Audit. BMJ open. 2018;8:e018444.
5. All Parliamentary Group for Continence Care. Continence Care Services: England 2013: Survey Report London; 2013.
6. Hunskaar S, Lose G, Sykes D, Voss S. The prevalence of urinary incontinence in women in four European countries. BJU International. 2004;93(3):324-30.
7. McGrowther C, Donaldson M, Shaw C, Matthews R, Hayward T, Dallosso H, et al. Storage symptoms of the bladder: prevalence, incidence and need for services in the UK. BJU International. 2004;93(6):763-9.
8. Thom D. Variation in estimates of urinary incontinence prevalence in the community: effects of differences in definition, population characteristics, and study type. Journal of the American Geriatrics Society. 1998;46(4):473-80.
9. Garry R, Fountain J, Mason S, Napp V, Brown J, Hawe J, et al. The eVALuate study: two parallel randomised trials, one comparing laparoscopic with abdominal hysterectomy, the other comparing laparoscopic with vaginal hysterectomy. Bmj. 2004;328:129.
10. Aarts J, Nieboer T, Johnson N, Tavender E, Garry R, Mol B, et al. Surgical approach to hysterectomy for benign gynaecological disease. The Cochrane database of systematic reviews. 2015;12(8):CD003677.
11. Darzi A. High quality care for all: NHS Next Stage Review final report. London: Department of Health; 2008.
12. Department of Health. The NHS Outcomes Framework 2016/17. London: Department of Health; 2016.
13. Mant J. Process versus outcome indicators in the assessment of quality of health care. Int J Qual Health Care. 2001;13:475-80.
14. Kiran A, Hilton P, Cromwell D. The risk of ureteric injury associated with hysterectomy: a 10-year retrospective cohort study. BJOG : an international journal of obstetrics and gynaecology. 2016;123(7):1184-91.
15. Olsen A, Smith V, Brgstrom J, Colling J, Clark A. Epidemiology of surgically managed pelvic organ prolapse and urinary incontinence. Obstetrics and gynecology. 1997;89(4):501-6.
16. Morling J, McAllister D, Agur W, Fischbacher C, Glazener C, Guerrero K, et al. Adverse events after first, single, mesh and non-mesh surgical procedures for stress urinary incontinence and pelvic organ prolapse in Scotland, 1997–2016: a population-based cohort study. The Lancet. 2017;389:629-40.
17. Keltie K, Elneil S, Monga A, Patrick H, Powell J, Campbell B, et al. Complications following vaginal mesh procedures for stress urinary incontinence: an 8 year study of 92,246 women. Scientific Reports. 2017;7(1):12015.
18. Geary RS, Knight HE, Carroll FE, Gurol-Urganci I, Morris E, Cromwell DA, et al. A step-wise approach to developing indicators to compare the performance of maternity units using hospital administrative data. BJOG: An International Journal of Obstetrics & Gynaecology.n/a-n/a.
19. Donabedian A. Quality assurance. Structure, process and outcome. Nursing Standard. 1992;11(Suppl QA):4-5.
20. NHS Digital. Hospital Episode Statistics 2017 [Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics.]
21. World Health Organization. International Classification of Diseases, 10th edition (ICD-10) [Available from: http://apps.who.int/classifications/icd10/browse/2016/en.]
22. Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures 4th revision (OPCS),. [Available from: https://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp]
23. Evans R, Zorlu G, Boseley S. Flaws in Hospital Episode Statistics revealed by FoI requests. The Guardian. 2010 10th June.

24. Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD. The accuracy of hospital discharge coding for hemorrhagic stroke. Acta neurologica Belgica. 2009;109(2):114-9.
25. Knight HE, Gurol-Urganci I, Mahmood TA, Templeton A, Richmond D, van der Meulen JH, et al. Evaluating maternity care using national administrative health datasets: how are statistics affected by the quality of data on method of delivery? BMC health services research. 2013;13:200.
26. Nouraei SA, O'Hanlon S, Butler CR, Hadovsky A, Donald E, Benjamin E, et al. A multidisciplinary audit of clinical coding accuracy in otolaryngology: financial, managerial and clinical governance considerations under payment-by-results. Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery. 2009;34(1):43-51.
27. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. J Public Health (Oxf). 2012;34(1):138-48.
28. Johal A, Mitchell D, Lees T, Cromwell D, van der Meulen J. Use of Hospital Episode Statistics to investigate abdominal aortic aneurysm surgery. The British journal of surgery. 2012;99(1):66-72.
29. Spiegelhalter D. Funnel plots for comparing institutional performance. Statistics in medicine. 2005;24(8):1185-202.
30. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Quality & safety in health care. 2005;14(5):347-51.
31. Royal College of Obstetricians and Gynaecologists. Recovering well: Information for you after a pelvic floor repair operation. In: Gynaecologists RCoOa, editor. London, UK2015.
32. Mulholland C, Harding N, Bradley S, Stevenson M. Regional variations in the utilization rate of vaginal and abdominal hysterectomies in the United Kingdom. Journal of Public Health Medicine. 1996;18(4):400-5.
33. Bottle A, Aylin P. Variations in vaginal and abdominal hysterectomy by region and trust in England. BJOG: An International Journal of Obstetrics and Gynaecology. 2005;112:326-8.
34. Brummer T, Jalkanen J, Heikkinen A, Kauko M, Mäkinen J, Seppälä T, et al. FINHYST, a prospective study of 5279 hysterectomies: complications and their risk factors. Human Reproduction. 2011;7:1741-51.
35. Dicker R, Greenspan J, Strauss L, Cowart M, Scally M, Peterson H, et al. Complications of abdominal and vaginal hysterectomy among women of reproductive age in the United States. The Collaborative Review of Sterilization. American journal of obstetrics and gynecology. 1982;144(7):841-8.
36. Pandit M, Alsop R. Using international data to set benchmarks for morbidity outcomes after hysterectomy. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics. 2016;133:84-8.
37. NHS Digital. Clinical Commissioning Group Outcomes Indicator Set (CCG OIS) 2017 [13/11/2017]. Available from: https://digital.nhs.uk/data-and-information/publications/ci-hub/ccg-outcomes-indicator-set.
38. NHS Digital. Commissioning Data Sets 2017 [Available from: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/commissioning-data-sets.]
39. Walker K, Neuberger J, Groene O, Cromwell D, van der Meulen J. Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. The Lancet. 2013;382:1674-7.
40. Mulley A. Improving productivity in the NHS. Bmj. 2010;341:c3965.
41. Burns E, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. Journal of Public Health. 2012;34(1):138-48.
42. CHKS. Payment by results data assurance framework: Key findings from the 2012/13 programmecapita. Birmingham; 2013.
43. Mowat A, Maher C, Ballard E. Surgical outcomes for low-volume vs high-volume surgeons in gynecology surgery: a systematic review and meta-analysis. American journal of obstetrics and gynecology. 2016;215(1):21-33.
44. Getting It Right First Time. Getting it right first time: Obstetrics and Gynaecology 2018 [Available from: http://gettingitrightfirsttime.co.uk/surgical-specialty/obstetrics-and-gynaecology/.]
45. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data resource profile: Hospital Episode Statistics Admitted Patient Care (HES APC). International Journal of Epidemiology. 2017;46(4):1093-i.
46. Department of Health. A simple guide to Payment by Results. Leeds: Department of Health; 2012.
47. Institute of Medicine (US). Crossing the quality chasm: a new health system for the 21st century. Washington (DC): National Academies Press; 2001.
48. Department of Health. NHS Outcomes Framework: at-a-glance. London, UK: Department of Health; 2016.
49. Shaw J, Taylor R, Dix K. Uses an abuses of performance data in healthcare. Dr Foster; 2015.