



Geary, RS; Knight, HE; Carroll, FE; Gurol-Urganci, I; Morris, E; Cromwell, DA; van der Meulen, JH (2017) A step-wise approach to developing indicators to compare the performance of maternity units using hospital administrative data. BJOG. ISSN 1470-0328 DOI: <https://doi.org/10.1111/1471-0528.15013>

Downloaded from: <http://researchonline.lshtm.ac.uk/4645567/>

DOI: [10.1111/1471-0528.15013](https://doi.org/10.1111/1471-0528.15013)

Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by-nc-nd/2.5/>

Article Type: Review Article

Article Category: General obstetrics

A step-wise approach to developing indicators to compare the performance of maternity units using hospital administrative data

Running title: How to develop maternity indicators from administrative data

Rebecca S Geary* (Corresponding author)

London School of Hygiene and Tropical Medicine, Department of Health Services Research and Policy, 15-17 Tavistock Place, London, WC1H 9SH, UK

Royal College of Obstetricians and Gynaecologists, Lindsay Stewart Centre for Audit and Clinical Informatics, 27 Sussex Place, London, NW1 4RG

Email: rebecca.geary@lshtm.ac.uk Tel. 02079272279

Hannah E Knight*

Royal College of Obstetricians and Gynaecologists, Lindsay Stewart Centre for Audit and Clinical Informatics, 27 Sussex Place, London, NW1 4RG

London School of Hygiene and Tropical Medicine, Department of Health Services Research and Policy, 15-17 Tavistock Place, London, WC1H 9SH, UK

***RSG and HEK contributed equally to this work and are therefore joint first authors**

Fran E Carroll

Royal College of Obstetricians and Gynaecologists, Lindsay Stewart Centre for Audit and Clinical Informatics, 27 Sussex Place, London, NW1 4RG

Ipek Gurol-Urganci

London School of Hygiene and Tropical Medicine, Department of Health Services Research and Policy, 15-17 Tavistock Place, London, WC1H 9SH, UK

Royal College of Obstetricians and Gynaecologists, Lindsay Stewart Centre for Audit and Clinical Informatics, 27 Sussex Place, London, NW1 4RG

Edward Morris

Royal College of Obstetricians and Gynaecologists, Lindsay Stewart Centre for Audit and Clinical Informatics, 27 Sussex Place, London, NW1 4RG

Norfolk and Norwich University Hospital (NNUH), Norwich NR4 7UZ

David A Cromwell

Royal College of Surgeons of England, Clinical Effectiveness Unit, 35-43 Lincoln's Inn Fields, London, WC2A 3PE, UK

London School of Hygiene and Tropical Medicine, Department of Health Services Research and Policy, 15-17 Tavistock Place, London, WC1H 9SH, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1471-0528.15013

This article is protected by copyright. All rights reserved.

Jan H van der Meulen

London School of Hygiene and Tropical Medicine, Department of Health Services Research and Policy, 15-17 Tavistock Place, London, WC1H 9SH, UK

Royal College of Obstetricians and Gynaecologists, Lindsay Stewart Centre for Audit and Clinical Informatics, 27 Sussex Place, London, NW1 4RG

ABSTRACT

Hospital administrative data are attractive for comparing performance of maternity units 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 including varying data quality, the limited detail on clinical risk factors and a lack of structural and user experience measures. This review illustrates how to develop performance indicators for maternity units using hospital administrative data, including methods to address the challenges administrative data pose.

Keywords: Maternity statistics, administrative data, HES, performance indicators

Tweetable abstract

How to develop maternity indicators from administrative data

BACKGROUND

There is growing interest in performance monitoring and quality improvement in healthcare in the UK and elsewhere.¹⁻³ Although quality of healthcare can be improved without measuring performance, for example, through educational programmes or clinical guidelines,^{4,5} accurate measurement of processes and outcomes of care is now seen as crucial for guiding service improvement.^{1,2,6,7}

In countries in which routinely collected hospital administrative datasets exist, these are attractive data sources for comparing performance of maternity units 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. A range of performance indicators derived from these data have been proposed for both primary and secondary care.^{8,9} However, despite ever-high levels of public interest in the safety and quality of maternity care, many countries lack robust, easily interpretable information on even basic maternal and perinatal outcomes¹⁰

In the UK, a number of initiatives have recently been introduced, aiming to improve this situation by drawing on routinely collected clinical and hospital administrative data to measure performance and variation in care between maternity units.¹¹⁻¹³ For example, the new National Maternity and Perinatal Audit in England, Scotland and Wales will link existing data sources, rather than introducing new bespoke data collections solely for the purposes of the audit.¹⁴ In Scandinavian countries, there is a long history of using medical birth registry databases, often further linked to other national databases, to conduct research into the organisation and outcomes of maternity care and to inform public health initiatives.¹⁵ Similar initiatives based on routine data linkage exist in Australia.¹⁶ A recent systematic review found that broader adoption of routine data linkage of perinatal health databases could yield substantial gains for research and surveillance.¹⁷ It therefore seems likely that, despite the inherent challenges of using these data, there will be a reliance on maternity indicators based, at least in part, on administrative data for some time to come.

Hospital administrative data have several advantages for describing care and outcomes. Where administrative data are readily available they are a cost-effective source of information. Where the majority of care is captured by these data, the risk of selection bias is reduced and sample sizes can be large. For example, in England >96% of all deliveries occur in NHS hospitals and are captured by administrative data (Hospital Episode Statistics, HES).¹⁸ Hospital administrative data also capture multiple procedures and diagnoses at the patient level, providing a rich description of patient characteristics and clinical risk factors. However, there are some important limitations of using

administrative data to develop performance indicators. There can be concerns about the accuracy and completeness of diagnosis and procedure coding,¹⁹ although there is mounting evidence that in England, most NHS trusts submit good quality data to HES.²⁰⁻²² Another limitation is that not all clinical information is captured in administrative data. Some risk factors such as BMI, smoking and alcohol consumption are often not recorded. This means they cannot be taken into account in risk-adjustment for case-mix, although record linkage can extend the range of data items available and thus can improve the validity and quality of routine data. In addition, administrative data lend themselves to process and outcome indicators; measures of structural and user experience are not normally available.

As a result of these challenges and opportunities, the use of hospital administrative data for performance monitoring requires caution and a robust methodology. However, information on how to derive maternity indicators from administrative data sources is lacking. Our aim is to address this by describing a transparent approach with explicit criteria. This approach can be used by those wanting to develop performance maternity indicators using HES data, the national administrative database of the English National Health Service (NHS), or administrative data available in other healthcare settings. Furthermore, the criteria can be used by clinicians to evaluate existing performance indicators.

This approach has been used to develop indicators for the Royal College of Obstetricians and Gynaecologists' (RCOG) Clinical Indicators Project and examples from this project will be used throughout.^(10, 11) The data source used for the examples is the HES database, containing records of admissions to NHS hospitals. Briefly, the HES database contains information on each episode of admitted patient care in the English NHS.²³ Each record contains data on patient demographics (such as age, sex and ethnicity), 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)²⁴ and procedures using the Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures, 4th revision (OPCS-4).²⁵ Each episode related to the delivery of a baby can also capture details about the labour and birth, such as parity, mode of delivery, gestational age and birthweight, in supplementary data fields known as the ‘maternity tail’. Each patient is assigned a unique identifier, allowing the study of longitudinal patterns of care or the number of previous births for a particular woman. For the example, delivery records were defined as those with information about a delivery in either the maternity tail or the OPCS fields in the financial year 2013/14.

DEVELOPING INDICATORS

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 consumers’ choice of healthcare provider.

Building on work carried out in several surgical specialties,(20) we assessed the suitability of using hospital administrative data for developing maternity indicators and developed a three-stage process: “identification”, “development and evaluation”, and “implementation and feedback” (summarised in Figure 1).^{10 11} The second stage involves evaluation against four criteria: “validity”, “statistical power”, “technical specification”, and “fairness”. Potential indicators must meet each criterion before being evaluated against the next.

Figure 1 about here

Seeking input from clinicians, methodologists and service users is a key characteristic of this indicator development process, and should be sought at each stage. Input can be via formal

consensus methods,²⁶ or less formal round-table discussions. To minimise bias, it is important that the discussions are facilitated by a neutral chair and to ensure that all key stakeholder groups are sufficiently represented. The specific stakeholder groups that are key may vary in different settings, depending on how maternity care is organised. At the outset, all stakeholders should receive a clear brief about the aims of the initiative, the process for indicator selection and the opportunities and limitations of the data source/s. The consensus group that guided the indicator development process described in this paper consisted of obstetricians and midwives active in the English NHS, health services researchers and lay women with recent experience of English maternity care (see Acknowledgments). Lay representatives were recruited from the RCOG's Women's Voices Involvement Panel.²⁷

Stage One: Indicator identification

To identify candidate indicators for development using administrative data, a valuable first step is a systematic review of the literature, including clinical guidelines. This also allows for an examination of associations between candidate indicators and important outcomes. Non-systematic approaches can be informative, but do not maximise the use of available evidence.⁸ In addition to identifying indicators from the literature, suggestions can be sought from stakeholders with an interest in measuring the performance of maternity services via surveys or face-to-face meetings, thereby reducing the potential impact of publication bias.

To provide a broad understanding of the performance and quality of a healthcare service as a whole it is important that a suite of indicators is "balanced". A balanced suite would ideally include indicators relating to the structure of care, the processes of care, or the outcomes of the care received⁶ throughout the care pathway, and including measures of user experience. Indicators derived from administrative data will tend to focus on process and outcome indicators as structural and user experience measures are not normally available in these datasets. However, they are

nonetheless important for understanding many care outcomes. Balancing the types of measures used to evaluate performance can also help to minimise the risk of indicators being taken out of context and in that way misinforming quality improvement initiatives.

Stage Two: Indicator development and evaluation

Given the challenges of using administrative data it is important to rigorously evaluate candidate indicators to address these issues as far as possible. In our process this entailed evaluation against four criteria in turn: “validity”, “statistical power”, “technical specification”, and “fairness”.

Criterion One: Validity

Clinical and lay input should be sought to identify which of the identified indicators are considered to be clinically meaningful, or in other words to, measure aspects of the service or the quality of care provided that are relevant to patients. For an indicator to be considered valid, it must also be likely that a difference in the indicator reflects a difference the quality of care, and a specific direction should reflect better quality. For example, a higher rate of “obstetric anal sphincter injury” can be thought to reflect poorer obstetric care. Indicators not meeting this criterion should be dropped at this stage. Examples of decisions to include, refine and exclude indicators based on assessments of validity in the RCOG Maternity Indicators Project are provided in Appendix S1.

A key consideration when using hospital administrative data to develop indicators of quality of care is whether denominators and numerators can be adequately captured. Once an indicator was identified as valid, input from clinicians was used to define the appropriate “denominator” (the group of patients for whom the indicator is relevant) and “numerator” (the state or the event of which the frequency is captured by the indicator). For example, an indicator reflecting the use of elective CS before 39 weeks without clinical indication would have as its denominator the number of patients who had an elective CS without a recorded clinical indication (e.g. gestational hypertension,

gestational diabetes, or poor fetal growth) and as its numerator the number of patients in this group who had an elective CS before 39 weeks.¹¹

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 hospitals may account for a large part of the observed variation. Conclusions about quality of care can only be reasonably drawn after differences due to these other factors are excluded. With criteria two, three and four we describe a transparent approach to address these issues in administrative datasets.

Criterion Two: Statistical power

An advantage of using administrative data to develop indicators is their often large sample size.

However, even in large datasets, where an event or a procedure that forms part of an indicator is rare, the statistical power to identify providers with truly poor performance is low. In this situation, no evidence of poor performance cannot be taken as evidence of acceptable performance.²⁹

Indicators should be reported at a level (clinic, hospital, NHS trust) which is appropriate to how care is commissioned and provided. However, where numbers within a unit are too small, a higher-level unit of analysis, or a longer timeframe should be considered. If this is not appropriate (given the way that care is commissioned or provided) the indicator cannot be judged to have met this statistical power criterion. We rejected some maternity indicators due to a small number of events per

hospital per year. For example, the maternal mortality rate in the UK is 8.5/100,000 pregnancies.³⁰

Therefore, the 'signal to noise' ratio for this measure is too low to detect true differences between hospitals.^{31 32} In this situation, composite indicators may be appropriate. For example, a composite maternal morbidity indicator has been proposed using Australian routine hospital data.^{33 34}

Criterion Three: Feasibility of technical specification

As administrative data are collected for administrative purposes rather than for research or quality improvement, not all data items required for specific indicators may be adequately captured. Valid and adequately powered indicators should therefore next be evaluated in terms of their technical specification. This comprises a detailed assessment of the available data source(s) to establish how well patient populations, important case-mix differences and the procedures or outcomes that define the indicator can be captured. In HES, this involves exploring the diagnosis (ICD-10) and procedure (OPCS-4) codes which can be used to define the indicator in preliminary analyses. The technical specification of the inclusions and exclusions defined in step 1 should also be evaluated.

Where data required to construct the indicators and identify the units of analysis are available, an assessment of the data quality and completeness should also be conducted. Identifying data quality issues that would affect our ability to define the appropriate populations for each indicator allowed us to be confident that indicators were based on data that met minimum standards. We propose assessing data quality overall, and by hospital, using three main methods:¹¹

- Investigation of the proportion of missing data
- Internal consistency between data items within HES^{21 35}. For example, we excluded hospitals in which less than 90% of the records had consistency of mode of delivery between the main HES record and the maternity tail (Appendix S1).
- Comparison with results from external studies

Examples of data quality assessments conducted as part of the Maternity Indicators Project are provided in Appendix S2. Full details of how data quality was assessed have been published elsewhere.¹¹

Another example of evaluating the technical specification of an indicator is a recent study that explored whether a composite maternal morbidity indicator developed using Australian routine data could be derived from HES data.³³ This study found that the quality of the relevant HES data meant that 11 conditions that were included in the Australian indicator would have to be excluded from the English indicator. These included eclampsia, obstetric embolism and cardiac arrest/failure, which are associated with increased risk of maternal mortality in the UK,³⁶ making the resultant indicator questionable in its ability to accurately estimate maternal morbidity during childbirth in England.

Criterion Four: Fairness

Patient characteristics may influence indications for procedures and treatments, as well as influencing outcomes. Indicators should only be used for comparative purposes where adequate adjustment has been made for key case-mix differences between populations of patients.

Calculating indicators without appropriate risk-adjustment may give rise to misleading results.³⁷⁻³⁹ A number of questions should be evaluated as part of robust risk-adjustment (Box 2).

Hospital administrative data are able to capture multiple procedures and diagnoses at an individual level, providing a rich description of the case-mix of patient characteristics and clinical risk factors.

However, not all clinical information is captured; risk factors such as BMI, smoking and alcohol consumption are not recorded, meaning they cannot be accounted for.

In maternity care, certain pregnancy characteristics can have a large impact on the care provided and on outcomes. To ensure that the indicators would allow fair comparisons among hospitals, we decided to focus on women with singleton, term, cephalic deliveries, whose maternity care is most affected by between-hospital and provider variation in clinical practices.^{40,41} Multiple births, preterm births and breech deliveries require very different management. Remaining differences in case-mix between hospitals can be addressed in several ways. First, indicators can be stratified by a

clinical condition that has a major influence on outcomes. For example, we stratified maternity indicators by parity (nulliparous and multiparous). Second, risk-adjustment using a regression model can be used. Further information on the methods and impact of risk adjustment in the RCOG's maternity indicators project is provided in Appendix S3. Identifying which factors to include in these risk adjustment models is complex, requiring knowledge of the relevant factors, statistical expertise and adequate data.

Following evaluation against these four criteria of: "validity", "statistical power", "technical specification", and "fairness" four in turn, 18 maternity indicators were developed from HES data (Table 1).¹¹

Table 1 about here

Stage Three: Implementation and feedback

Given high levels of public interest in the quality of maternity care, and the challenges associated with using administrative data for this purpose, the use of indicators for performance assessment derived from administrative data needs to be cautiously implemented. A feasibility phase in which hospital-specific results are published anonymously can generate buy-in from those who will ultimately use the indicators. Individual, personalised feedback of results to the hospitals may also give them an opportunity to address identified data quality issues.^{10 11} Also, careful consideration should be given to the best methods for reporting results to achieve maximum impact with the intended audiences.^{43 44} This may be static reports, interactive online formats, and/or a series of local or regional discussion meetings. Finally, it is important to encourage those who use indicators not to interpret the results of individual indicators in isolation, but to look at a suite as a whole, considering possible relationships between indicators.⁴⁵ For example, in maternity care there is an association between lower pre-labour CS rates and therefore higher vaginal delivery rates on the

one hand but also higher emergency CS rates, which in turn can influence outcomes such as length of stay or readmission post-delivery.

CONCLUDING REMARKS

There is international interest in using indicators derived from administrative data to drive improvements in maternity care^{3 16} but information on how to derive indicators from administrative data, addressing the challenges presented by these data, is lacking. We present an approach for developing indicators using administrative data that has been well received by healthcare professionals and addresses many of the challenges of using administrative data for this purpose.^{10 11}

Key features of this process are explicit data quality checks, risk-adjusting hospital results for differences in patient case-mix, and clinical and lay input at all stages. Some of the indicators developed have already been incorporated into local monitoring systems and national outcome frameworks (Box 1) and will be developed further as part of the new National Maternity and Perinatal Audit.¹⁴ Our indicator development process is also more “streamlined” than others,^{3 46} with three steps and four evaluation criteria. This supportive approach also included collaboration between those developing indicators and those whose performance they are designed to monitor.

Box 1 about here

Overall, hospital administrative data sources are attractive for comparing performance of maternity units within and between countries due to their often large sample sizes, lack of selection bias, and the relatively low costs of accessing these data compared to those of conducting primary data collection. However, using administrative data to develop maternity indicators also present challenges, and the development of the indicators described in this review has triggered debate about the use of administrative data for this purpose. Such debate has included concerns about the accuracy and completeness of coding, that data quality may vary between healthcare providers, the

lack of detail on risk factors such as BMI, smoking and alcohol consumption and the absence of structural and user experience measures.^{19,47} However, a systematic review of discharge coding accuracy in administrative UK data found that primary diagnosis accuracy improved from 73.8% to 96.0% in the last decade, concluding that administrative data are sufficiently robust to use for research and managerial decision-making.⁴⁸ For each indicator proposed, careful evaluation of how well it can be derived from administrative data to meet “validity”, “statistical power”, “technical specification” and “fairness criteria” allows some of these challenges to be addressed. This transparent approach to developing indicators using administrative data could also be applied in other specialties, for primary care, and for international, national or regional comparisons.

Hospital administrative data are not the perfect data source for developing indicators of maternity care quality. Because administrative data are not collected for research purposes the performance indicators available may differ from core outcome sets used in clinical trials that focus on quality improvement and safety. The development of more clinically detailed routine maternity datasets⁴⁹ will ultimately allow for improvement of existing indicators, and the development of new indicators, producing a more balanced picture of the quality of maternity care. However, until centrally-available electronic maternity records become the norm, routine hospital administrative data, linked with other sources of clinical and user experience data where possible, will be the key data source for performance indicators. Some countries, such as Denmark, Finland, Norway and Sweden, are ahead of the game in integrating data linkage into their routine perinatal health surveillance systems and making these data available for research, but this is not a universal practice even in high-income countries with access to electronic hospital administrative data.¹⁷ Standardisation of performance measures derived from administrative data research would be desirable to facilitate comparisons both nationally and internationally.

Accepted Article

It seems likely that there will be a reliance on maternity indicators based, at least in part, on administrative data for some time to come. In light of this, methods for addressing the challenges posed by administrative data for the development of performance indicators are sorely needed. The transparent approach detailed in this paper aims to contribute to this effort. Our approach has led to the development of maternity indicators that have been adopted at a local and national-level, and addresses many of the issues raised about the usefulness of administrative data for performance monitoring.

Box 1: examples of use of indicator definitions or data

Information from the maternity indicators project has already been used by trusts and incorporated into local monitoring systems, national outcomes frameworks. For example:

- The definition of the indicator “Elective caesarean section without indication before 39 weeks of gestation” has been proposed for the Clinical Commissioning Group Outcomes Indicator Set
- Several Clinical Networks held regional workshops to encourage trusts within the same region to compare results and reflect on the cause of any differences in practices or outcomes
- A number of indicators from this project have been included in regional dashboards.⁵⁰
- The indicators have been used by trusts in the following ways (based on a small evaluation survey carried out in May 2016; n=19 trusts):
 - Discussed with clinical board/Senior Management Team: 55% (n=11)
 - Led to an internal audit: 30% (n=6)
 - Led to an investigation of data collection/coding/provision: 45% (n=9)
 - Led to a change in data systems/ability to provide data in the future: 25% (n=5)

Case Mix

How big are the important case-mix differences between hospitals/trusts?

Data

Does sufficient detail on case-mix exist in the available data?

If not, could data linkage be used to obtain these data from other sources?

Unmeasured Confounding

Which factors do not have data available which could result in unmeasured confounding?

Impact of Adjustment

What is the impact of risk-adjustment on the differences between the hospitals/trusts?

Box 2: Risk Adjustment: Questions to Ask**Ethical approval**

The study is exempt from UK National Research Ethics Service (NRES) approval because it involved the analysis of an existing dataset of anonymized data for service evaluation. HES data were made available by NHS Digital (Copyright 2015, Re-used with the permission of NHS Digital. All rights reserved.) Approvals for the use of anonymized HES data were obtained as part of the standard NHS Digital data access process.

Funding

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 Royal College of Obstetricians and Gynaecologists.

Contributors

RSG, HEK, IGU, EM and JvdM conceived the review. FC, HEK and IGU performed the data analysis for the Clinical Indicators Project that informed this review, with technical support from DAC and advice from JvdM and EM. All authors contributed to the writing of the manuscript.

Disclosure of interests

The authors declare that they have no conflicts of interest. The ICMJE disclosure forms are available as online supporting information.

Acknowledgments

We would like to thank the Clinical Effectiveness Unit (CEU) at the Royal College of Surgeons of England for supporting this project and providing access to the Hospital Episode Statistics (HES) data used to develop the maternity indicators. In particular, we are grateful to Lynn Copley for her support with data extraction. We are indebted to the members of the consensus group who guided the development and specification of the maternity indicators described here as well as the interpretation of the findings(11). We would also like to thank the hospitals that contacted us with their feedback during the course of this project.

References

1. Department of Health. The NHS Outcomes Framework 2016/17. London: Department of Health. , 2016.
2. Darzi A. High quality care for all: NHS Next Stage Review final report. . London: Crown, 2008.
3. AHRQ. AHRQ Quality Indicators. Guide to Patient Safety Indicators.: Department of Health and Human Services Agency for Healthcare Research and Quality, 2003.
4. Forrest D, Hoskins A, Hussey R. Clinical guidelines and their implementation. *Postgrad Med J* 1996;**72**(843):19-22.
5. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;**342**(8883):1317-22.
6. Donabedian A. Quality assurance. Structure, process and outcome. *Nurs Stand* 1992;**7**(11 Suppl QA):4-5.
7. Irvine EJ, Zhou Q, Thompson AK. The Short Inflammatory Bowel Disease Questionnaire: a quality of life instrument for community physicians managing inflammatory bowel disease. CCRPT Investigators. Canadian Crohn's Relapse Prevention Trial. *Am J Gastroenterol* 1996;**91**(8):1571-8.
8. Campbell SM, Braspenning J, Hutchinson A, et al. Research methods used in developing and applying quality indicators in primary care. *Quality & safety in health care* 2002;**11**(4):358-64.
9. Powell AE, Davies HT, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Quality & safety in health care* 2003;**12**(2):122-8.
10. RCOG. Patterns of Maternity Care in English NHS Hospitals. London: RCOG, 2013.

11. RCOG. Patterns of Maternity Care in English NHS Trusts. London: RCOG, 2016.
12. NHS Digital. Maternity Services Monthly Statistics - October, 2016, Experimental. Secondary Maternity Services Monthly Statistics - October, 2016, Experimental 2017. <http://content.digital.nhs.uk/maternityandchildren/maternityreports>.
13. Which? Birthchoice. Secondary 2017. <http://www.which.co.uk/birth-choice/>.
14. National Maternity and Perinatal Audit. Secondary 2017. <http://www.maternityaudit.org.uk/>.
15. Socialstyrelsen. The Swedish Medical Birth Register. Secondary The Swedish Medical Birth Register. <http://www.socialstyrelsen.se/register/halsodataregister/medicinskafodelseregistret/inenglish>.
16. Australian Institute of Health and Welfare. National Core Maternity Indicators. Secondary National Core Maternity Indicators. <http://www.aihw.gov.au/ncmi/>.
17. Delnord M, Szamotulska K, Hindori-Mohangoo AD, et al. Linking databases on perinatal health: a review of the literature and current practices in Europe. *Eur J Public Health* 2016;**26**(3):422-30.
18. Birthplace in England Collaborative G, Brocklehurst P, Hardy P, et al. Perinatal and maternal outcomes by planned place of birth for healthy women with low risk pregnancies: the Birthplace in England national prospective cohort study. *Bmj* 2011;**343**:d7400.
19. Evans R, Zorlu G, Boseley S. Flaws in Hospital Episode Statistics revealed by FoI requests. *The Guardian* 2010 10th June.
20. Kirkman MA, Mahattanakul W, Gregson BA, et al. The accuracy of hospital discharge coding for hemorrhagic stroke. *Acta neurologica Belgica* 2009;**109**(2):114-9.
21. Knight HE, Gurol-Urganci I, Mahmood TA, 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.
22. Nouraei SA, O'Hanlon S, Butler CR, 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.
23. NHS Digital. Hospital Episode Statistics. Secondary Hospital Episode Statistics. <http://content.digital.nhs.uk/hes>.
24. World Health Organization. International Classification of Diseases, 10th edition (ICD-10). Secondary International Classification of Diseases, 10th edition (ICD-10). <http://apps.who.int/classifications/icd10/browse/2016/en>.
25. Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures, 4th revision (OPCS). Secondary Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures, 4th revision (OPCS). <http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4>.
26. Black N, Murphy M, Lamping D, et al. Consensus development methods: a review of best practice in creating clinical guidelines. *J Health Serv Res Policy* 1999;**4**(4):236-48.
27. RCOG Women's Voices Involvement Panel. Secondary 2017. <https://www.rcog.org.uk/en/patients/womens-voices-involvement-panel/>.
28. NICE. CG132: Caesarean section (2011). Secondary CG132: Caesarean section (2011). <https://www.nice.org.uk/guidance/cg132>.
29. Walker K, Neuburger J, Groene O, et al. Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. *Lancet* 2013;**382**(9905):1674-7.
30. Manktelow BM, Smith LK, Evans TA, Hyman-Taylor P, Kurinczuk JJ, Field DJ, et al. Perinatal Mortality Surveillance Report UK Perinatal Deaths for births from January to December 2013 Leicester: The Infant Mortality and Morbidity Group, Department of Health Sciences, University of Leicester, 2015.
31. Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process

- and outcome in treating acute myocardial infarction. *Bmj* 1995;**311**(7008):793-6.
32. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *Jama* 2001;**286**(4):415-20.
 33. Nair M, Kurinczuk JJ, Knight M. Establishing a National Maternal Morbidity Outcome Indicator in England: A Population-Based Study Using Routine Hospital Data. *PLoS One* 2016;**11**(4):e0153370.
 34. Roberts CL, Cameron CA, Bell JC, et al. Measuring maternal morbidity in routinely collected health data: development and validation of a maternal morbidity outcome indicator. *Medical care* 2008;**46**(8):786-94.
 35. Cromwell DA, Knight HE, Gurol-Urganci I. Parity derived for pregnant women using historical administrative hospital data: accuracy varied among patient groups. *Journal of clinical epidemiology* 2014;**67**(5):578-85.
 36. Knight M, Tuffnell D, Kenyon S, et al. Saving Lives, Improving Mothers' Care - Surveillance of maternal deaths in the UK 2011-13 and lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2009-13. Oxford: National Perinatal Epidemiology Unit, University of Oxford 2015.
 37. Korst LM, Gornbein JA, Gregory KD. Rethinking the cesarean rate: how pregnancy complications may affect interhospital comparisons. *Medical care* 2005;**43**(3):237-45.
 38. Bailit JL, Grobman WA, Rice MM, et al. Risk-adjusted models for adverse obstetric outcomes and variation in risk-adjusted outcomes across hospitals. *American journal of obstetrics and gynecology* 2013;**209**(5):446 e1-46 e30.
 39. Pasternak DP, Pine M, Nolan K, et al. Risk-adjusted measurement of primary cesarean sections: reliable assessment of the quality of obstetrical services. *Qual Manag Health Care* 1999;**8**(1):47-54.
 40. Robson MS, Scudamore IW, Walsh SM. Using the medical audit cycle to reduce cesarean section rates. *American journal of obstetrics and gynecology* 1996;**174**(1 Pt 1):199-205.
 41. Main EK, Moore D, Farrell B, et al. Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. *American journal of obstetrics and gynecology* 2006;**194**(6):1644-51; discussion 51-2.
 42. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Statistics in medicine* 2005;**24**(8):1185-202.
 43. Hendricks M. *Handbook of Practical Program Evaluation. Chapter 23. Making a Splash: Reporting Evaluation Results Effectively*, 1994.
 44. Centers for Disease Control and Prevention. Evaluation Reporting: A Guide to Help Ensure Use of Evaluation Findings. Atlanta, GA: US Dept of Health and Human Services, 2013.
 45. Pyykonen A, Gissler M, Jakobsson M, et al. Determining obstetric patient safety indicators: the differences in neonatal outcome measures between different-sized delivery units. *BJOG : an international journal of obstetrics and gynaecology* 2014;**121**(4):430-7.
 46. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. *Int J Qual Health Care* 2003;**15 Suppl 1**:i5-11.
 47. Chappell LC, Calderwood C, Kenyon S, et al. Understanding patterns in maternity care in the NHS and getting it right. *Bmj* 2013;**346**:f2812.
 48. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *J Public Health (Oxf)* 2012;**34**(1):138-48.
 49. Health and Social Care Information Centre. Maternity Services Dataset. Secondary Maternity Services Dataset. <http://www.hscic.gov.uk/maternityandchildren/maternity>.
 50. South East Coast Strategic Clinical Networks. Maternity Dashboard. Secondary Maternity Dashboard. <http://maternitydashboard.secsn.nhs.uk/index.php/definitions-list>.

Figure Titles and Legends

Figure 1: How to develop relevant, rigorous and robust clinical indicators

Table 1: Indicators developed from HES for the RCOG's Maternity Indicators Project

1. UNASSISTED VAGINAL DELIVERIES	Population subset
1a) Proportion of spontaneous, unassisted vaginal deliveries	Primip/Multip
2. INDICATORS RELATED TO INDUCTION OF LABOUR	
2a) Proportion of induced labours	Primip/Multip
2b) Proportion of induced labours in deliveries between 37 and 39 weeks of gestation	Primip/Multip
2c) Proportion of induced labours in deliveries ≥ 42 weeks of gestation	Primip/Multip
3. INDICATORS RELATING TO CAESAREAN SECTION	
3a) Proportion of deliveries by caesarean section	Primip/Multip
3b) Proportion of induced labours resulting in emergency caesarean section	Primip/Multip
3c) Proportion of spontaneous labours resulting emergency caesarean section	Primip/Multip
3d) Proportion of prelabour caesarean sections	Primip/Multip
3e) Proportion of prelabour caesarean sections performed before 39 weeks of gestation without clinical indication	Pre
3f) Proportion of vaginal births following a primary caesarean section (VBAC)	Multip
4. INVOLVEMENT OF INSTRUMENTS	
4a) Proportion of deliveries involving instruments	Primip/Multip
5. EPISIOTOMY	
5a) Proportion of episiotomies among vaginal deliveries	Primip/Multip
5b) Proportion of episiotomies among instrumental deliveries	F/Va
6. INDICATORS RELATING TO 3rd AND 4th DEGREE TEARS	
6a) Proportion of third- and fourth-degree perineal tears among vaginal deliveries	Primip/Multip
6b) Proportion of third- and fourth-degree perineal tears among unassisted vaginal deliveries	Primip/Multip
6c) Proportion of third- and fourth-degree perineal tears among assisted vaginal deliveries	Primip/Multip
7. ADMISSIONS TO HOSPITAL FOLLOWING DELIVERY	
7a) Unplanned maternal readmission to hospital within 42 days of delivery	V/CS
7b) Unplanned neonatal readmission to hospital within 28 days of birth	NB
<i>Footnote: For all indicators, multiple and preterm deliveries were excluded. Women who delivered a baby with a non-cephalic presentation were also excluded, apart from for indicators 3e and 7b. Primip=Primiparous; Multip=Multiparous; CS=Caesarean section deliveries; F=Forceps, NB= Normal Birthweight Infants; Pre= Subset of prelabour caesarean section deliveries including women with non-cephalic presentation OR where 1 or 2 previous caesarean sections; Va=Vacuum</i>	

HOW TO DEVELOP CLINICAL INDICATORS

1: IDENTIFICATION

Identify candidate indicators with a systematic literature review

Use clinical and lay input to decide which are important

2: DEVELOPMENT AND EVALUATION

Evaluate indicators in terms of validity, statistical power, technical specification and fairness in turn.

If an indicator fails to meet one criteria, remove it

2A: VALIDITY

Are differences in an indicator likely to reflect quality of care?

2B: STATISTICAL POWER

What is the chance of detecting a true outlier?

2C: TECHNICAL SPEC

How well can the data define patients, procedures or outcomes for an indicator?

2D: FAIRNESS

How different are patients treated by different units? Can we capture & adjust for this?

3) IMPLEMENTATION AND FEEDBACK

- Engage users early
- Publish indicators anonymously first to build trust and drive up data quality
- Interpret related indicators together, not alone