Appendix S1: Supplementary Methods

Sampling

The sample size was determined pragmatically in order to obtain a full year's data at two centres to even out any seasonal variation and increase generalisability. However, as the total number of women needed for this approach was too large to feasibly assess all their handheld notes individually within available resources. We constructed an achievable stratified sampling plan to maximise the power of the study within these resources; this led to a total weighted sample size of 1897 women, from the 10213 who delivered at the two centres.

A weighted sample[1,2] of records was selected by two researchers within a week of importing the study data to provide a representative sample set. Paper records were randomly selected for detailed review on the basis of the estimated blood loss (EBL) recorded in the hospital database, as follows: EBL>25-499ml, one in twelve records; EBL 500-999ml, one in six records; EBL≥1000ml, EBL<25ml, no recorded EBL or blood/blood products ordered, all records. Random selection was achieved by witnessed and recorded paired dice throws. By selecting postpartum haemorrhage (PPH) subjects on the basis of blood loss category together with a random selection of unaffected women, high statistical power was maintained. Standard errors were adjusted for the sampling strategy using the Huber-White estimator[3].

This sampling structure was corrected for in the analysis by using sampling weights. Except when discussing errors in the electronic data, all results refer to the corrected blood loss values based on the reviewed handheld notes.

The process of random selection, using fair dice, was validated by testing for any systematic bias towards any particular number. The frequencies of the first 2275 throws were 371, 372, 396, 369, 383, 384 for each value; Chi-sq = 1.432, P = .921 (5 degrees of freedom).

Inaccuracies, missing data and exclusions

There were two small groups of women for whom the blood loss as documented in the notes may have been an underestimate. In both cases, they were categorised as <500ml following the protocol even though features suggested a PPH: 1) those with a fall in haemoglobin of ≥4 g/l between late pregnancy (≥34 weeks') and postnatal (n=7); 2) those where clinicians undertook actions indicative of PPH (n=39); as there are other possible explanations for these features.

Some subjects with important routine data missing were excluded from analyses that used it: this included 13 women who had no valid English postcode to which deprivation scores could be matched and 16 women without recorded body mass index (BMI).

The largest source of missing data was intrapartum temperature (523 women, 38% of the sample, 27% after allowing for weighting). The choices were to exclude missing data, impute normal temperature to missing cases or treat 'unknown temperature' as a separate category. On the basis that all but 2 women undergoing elective caesarean section (CS) had no temperature measurement recorded, exclusion would have biased the results, whereas imputation as normal would have attenuated the real impact. For the main analysis, missing values were regarded as a separate category. To check the

importance of this, the analysis was repeated omitting all women with missing values. This made no substantive change, with an identical pattern of significance levels. Unrecorded timing of rupture of membranes (181 women) and unrecorded suture interval following vaginal delivery (143 women) were treated similarly.

By the process of logistic regression, perfect predictors are excluded automatically: 26 subjects with either major or anterior placenta praevia all had PPH, as did 10 women with evidence of chorioamnionitis.

Pre-defined grouping of risk factors for PPH

a) *Pre-pregnancy:* i) socio-demographic (age, ethnicity), ii) deprivation (index of multiple local deprivation and seven subscales[4] including barriers to housing and services; crime and disorder; education, skills and training; employment; health, deprivation and disability; income; and living environment), iii) general & medical risk factors, iv) obstetric history (previous PPH, previous CS, parity).

b) *Pregnancy:* v) current pregnancy, vi) antenatal care (first antenatal appointment data, antenatal day unit attendances and hospital admissions subdivided by reason for admisson), vii) placenta praevia, viii) antepartum haemorrhage (APH) and urinary tract infection (UTI), ix) pre-eclampsia and anaemia (gestational hypertension, pre-eclampsia, polyhydramnios, anaemia), and x) medications pre-delivery.

c) *Labour and birth:*

xi) gestation at birth, xii) birthweight, xiii) onset of labour, xiv) intrapartum factors (duration of rupture of membranes, Prostin use, temperature, chorioamnionitis, Synctocinon® use, regional analgesia and anaesthesia), xv) birth (mode of delivery), xvi), third stage (management, retained placenta, interval to suturing).

For each of these, a series of models was fitted, in which the role of the elements of each group of potential predictors were considered, adjusting for the influence of predictors identified from previous groups (potential confounders). By fitting new elements singly, factors not significantly related to outcome were removed. In this way, adjustment was made for potential confounders at each stage[5]. This method ensured that useful information for prediction of PPH or progression was captured at each of the 16 stages wherever possible. Inclusion of all variables in a single selection model would drive out important early predictors that were mediated through later events.

For three variables defined in multiple ways (intrapartum temperature, placenta praevia and APH), alternative models were fitted to better characterise their influence. Only differences in temperature above 37.0 C were related to outcome, and all low and normal temperatures were coded together. Only major and anterior placenta praevia were related to outcome. Minor posterior placenta praevia was not a risk factor for PPH. APH linked to serious symptoms (recurrent at least 3 episodes, like a period or heavier) or where clinicians diagnosed a placental cause were considered separately from other APHs.

Bootstrapping

A bootstrap process to repeat the complex procedures detailed above on randomly different versions of the dataset was considered but not undertaken because it would add another layer of statistical complexity to a study that is already extremely difficult to report concisely and clearly.

Multiple testing

Multiple regression models use all relevant predictors significant in univariate analysis. Significant and non-significant results are reported.

No formal adjustment was made for multiple testing[3,5,6] as a single primary endpoint was not relied upon; the totality of the evidence from a number of related sources was considered. In summaries and discussion, greatest emphasis was put on the most clinically important and statistically significant results. All principal conclusions were supported by a number of related tests, usually at levels beyond $p<0.001$. Variables were predictive when considered alone, when adjusting for potential confounders from earlier groups, and in up to three separate, but related regression models.

Additional References

1. Korn, E.L., Graubard, B.I. (1999) Analysis of Health Surveys. Wiley, New York.
2. Malgarini, M. (2005) Efficient sample design and weighting methodologies. Joint European Commission-OECD workshops on international development of business and consumer tendency surveys, Task Force on Harmonisation of Survey Operation and Technical Design.
3. Huber, P.J. (1967) Behaviours of likelihood estimates under non-standard conditions. University of California Press, California.
4. McLennan, D., Barnes, H., Noble, M., Davies, J., Garratt, E., Dibben, C. (2011) The English Indices of Deprivation 2010. Department for Communities and Local Government.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6871/1871208.pdf accessed 16th August 2013

5. Hernan, M.A., Hernamdez-Diaz, S., Werler, M.M., Mitchell, A.A. (2002) Causal Knowledge as a Prerequisite for Confounding Evaluation: An Application to Birth Defects Epidemiology. Am J Epidemiol 155 (2),176-84.
6. Rothman, K.J. (1990) No adjustments needed for multiple comparisons. Epidemiol 1, 43-6.