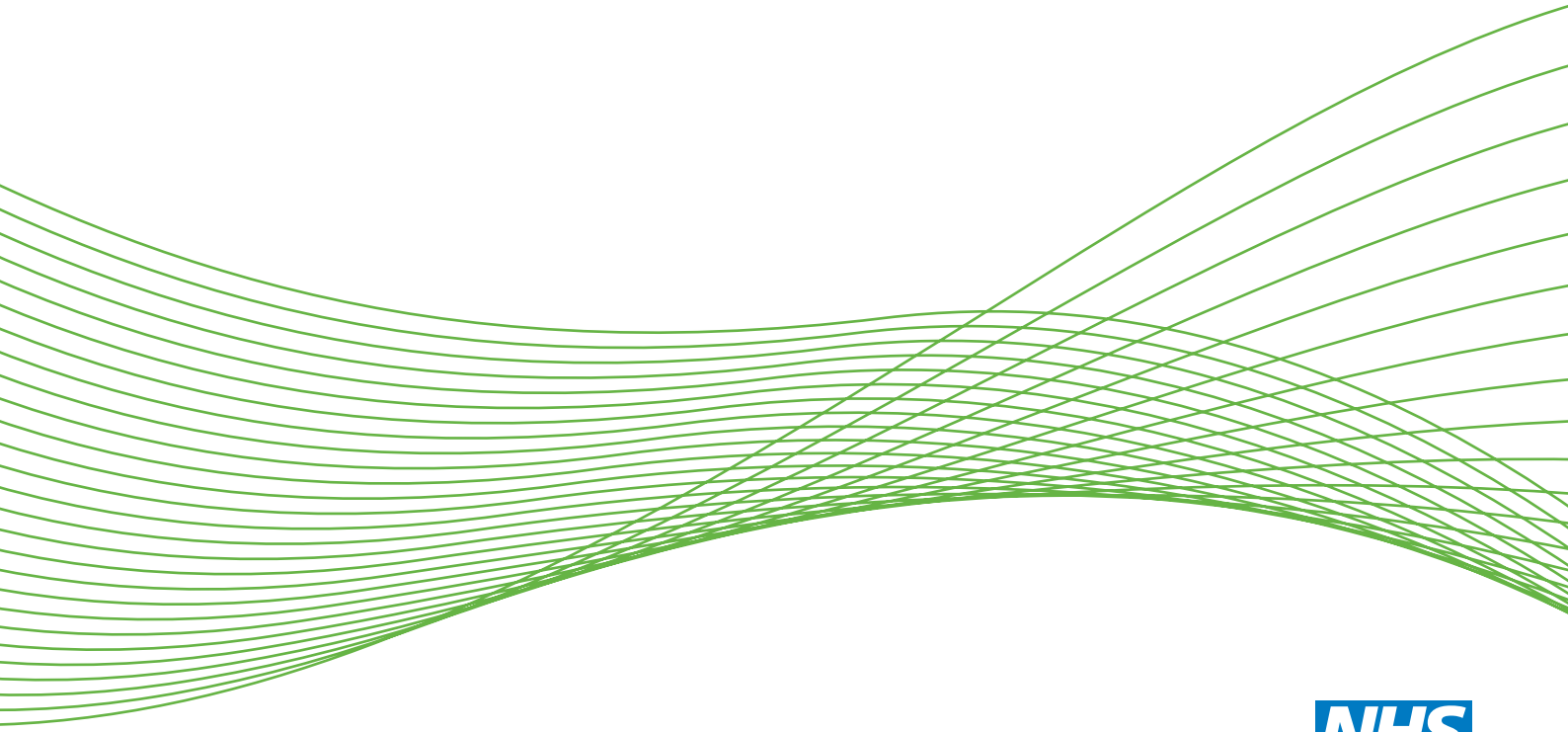
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 19 ISSUE 22 MARCH 2015
ISSN 1366-5278

A study of cellular counting to determine minimum thresholds for adequacy for liquid-based cervical cytology using a survey and counting protocol

Henry C Kitchener, Matthew Gittins, Mina Desai, John HF Smith, Gary Cook, Chris Roberts and Lesley Turnbull



National Institute for Health Research

DOI 10.3310/hta19220

A study of cellular counting to determine minimum thresholds for adequacy for liquid-based cervical cytology using a survey and counting protocol

Henry C Kitchener,[1]* Matthew Gittins,[2] Mina Desai,[3] John HF Smith,[4] Gary Cook,[5] Chris Roberts[2] and Lesley Turnbull[6]

[1]Institute of Cancer Sciences, St. Mary's Hospital, University of Manchester, Manchester, UK
[2]Institute of Population Health, University of Manchester, Manchester, UK
[3]Cytology Department, Clinical Sciences Building, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK
[4]Department of Histology and Cytology, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[5]Stepping Hill Hospital, Stockport NHS Foundation Trust, Stockport, UK
[6]Liverpool Women's Hospital, Liverpool, UK

*Corresponding author

Declared competing interests of authors: Dr John HF Smith reports grants from the Health Technology Assessment programme, during the conduct of the study; personal fees, accommodation and travel expenses from BD (Becton Dickinson) Europe Speaker Bureau, BD Asia-Pacific Speaker Bureau, outside the submitted work. Dr Mina Desai has received travel money and accommodation paid for by BD company to lecture on the Scientific Symposium in Sweden and India. Professor Henry C Kitchener is the chairperson of the Advisory Committee for Cervical Screening, but all views are reported here are those of the author and not of Public Health England.

Published March 2015
DOI: 10.3310/hta19220

This report should be referenced as follows:

Kitchener H, Gittins M, Desai M, Smith JHF, Cook G, Roberts C, *et al*. A study of cellular counting to determine minimum thresholds for adequacy for liquid-based cervical cytology using a survey and counting protocol. *Health Technol Assess* 2015;**19**(22).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 05/41/02. The contractual start date was in October 2007. The draft report began editorial review in May 2014 and was accepted for publication in August 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

A study of cellular counting to determine minimum thresholds for adequacy for liquid-based cervical cytology using a survey and counting protocol

Henry C Kitchener,[1]* Matthew Gittins,[2] Mina Desai,[3] John HF Smith,[4] Gary Cook,[5] Chris Roberts[2] and Lesley Turnbull[6]

[1]Institute of Cancer Sciences, St. Mary's Hospital, University of Manchester, Manchester, UK
[2]Institute of Population Health, University of Manchester, Manchester, UK
[3]Cytology Department, Clinical Sciences Building, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK
[4]Department of Histology and Cytology, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[5]Stepping Hill Hospital, Stockport NHS Foundation Trust, Stockport, UK
[6]Liverpool Women's Hospital, Liverpool, UK

*Corresponding author henry.kitchener@manchester.ac.uk

Background: Liquid-based cytology (LBC) for cervical screening would benefit from laboratory practice guidelines that define specimen adequacy for reporting of slides. The evidence base required to define cell adequacy should incorporate both ThinPrep™ (TP; Hologic, Inc., Bedford, MA, USA) and SurePath™ (SP; BD Diagnostics, Burlington, NC, USA), the two LBC systems used in the UK cervical screening programmes.

Objectives: The objectives of this study were to determine (1) current practice for reporting LBC in England, Wales and Scotland, (2) a reproducible method for cell counting, (3) the cellularity of slides classified as inadequate, negative or abnormal and (4) the impact of varying cellularity on the likelihood of detecting cytological abnormalities.

Design: The study involved four separate arms to pursue each of the four objectives. (1) A questionnaire survey of laboratories was conducted. (2) A standard counting protocol was developed and used by three experienced cytopathologists to determine a reliable and reproducible cell counting method. (3) Slide sets which included a range of cytological abnormalities were each sent to three laboratories for cell counting to study the correlation between cell counts and reported cytological outcomes. (4) Dilution of LBC samples by fluid only (unmixed) or by dilution with a sample containing normal cells (mixed) was performed to study the impact on reporting of reducing either the total cell count or the relative proportion of abnormal to normal cells.

Setting: The study was conducted within the cervical screening programmes in England, Wales and Scotland, using routinely obtained cervical screening samples, and in 56 participating NHS cervical cytology laboratories.

Participants: The study involved only routinely obtained cervical screening samples.

Interventions: There was no clinical intervention.

Main outcome measures: The main outcome measures were (1) reliability of counting method, (2) correlation of reported cytology grades with cellularity and (3) levels of detection of abnormal cells in progressively diluted cervical samples.

Results: Laboratory practice varied in terms of threshold of cellular adequacy and of morphological markers of adequacy. While SP laboratories generally used a minimum acceptable cell count (MACC) of 15,000, the MACC employed by TP laboratories varied between 5000 and 15,000. The cell counting study showed that a standard protocol achieved moderate to strong inter-rater reproducibility. Analysis of slide reporting from laboratories revealed that a large proportion of the samples reported as inadequate had cell counts above a threshold of 15,000 for SP, and 5000 and 10,000 for TP. Inter-rater unanimity was greater among more cellular preparations. Dilution studies demonstrated greater detection of abnormalities in slides with counts above the MACC and among slides with more than 25 dyskaryotic cells.

Conclusions: Variation in laboratory practice demonstrates a requirement for evidence-based standards for designating a MACC. This study has indicated that a MACC of 15,000 and 5000 for SP and TP, respectively, achieves a balance in terms of maintaining sensitivity and low inadequacy rates.

Future work: The findings of this study should inform the development of laboratory practice guidelines.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of tables

List of figures

List of abbreviations

ANOVA	analysis of variance
ARTISTIC	A Randomised Trial In Screening To Improve Cytology
CI	confidence interval
CIN	cervical intraepithelial neoplasia
GEE	generalised estimating equation
HPV	human papillomavirus
ICC	intraclass correlation coefficient
LBC	liquid-based cytology
MACC	minimum acceptable cell count
NHSCSP	National Health Service Cervical Screening Programme
NICE	National Institute for Health and Care Excellence
OR	odds ratio
SD	standard deviation
SOP	standard operating procedure
SP	SurePath™
TBS	the Bethesda System
TP	ThinPrep™

Plain English summary

The introduction of liquid-based cytology, using the commercial SurePath™ (SP; BD Diagnostics, Burlington, NC, USA) and ThinPrep™ (TP; Hologic, Inc., Bedford, MA, USA) systems, to the UK Cervical Screening Programme has resulted in the proportion of inadequate slides falling from 7–8% to 1–2%. There is uncertainty regarding the minimum number of cells needed within these preparations to provide a reliable reading because an important reason for inadequate slides is insufficient cells.

This study, which was performed between 2008 and 2011 using routinely obtained cervical screening samples, sought to address this uncertainty by means of (1) surveying current laboratory practice; (2) assessing the reliability (between experienced readers) of counting the number of cells on a slide; (3) evaluating the relationship between cell counts and the grade of cellular abnormalities reported across a range of laboratories; and (4) evaluating the effect of cell dilution on the reliability of reporting.

The participating laboratories reported variable practice in defining an adequate cell count and cell counting protocol. When a pre-specified cell counting protocol was adhered to, counting was moderately/strongly reproducible. The currently reported 'inadequate' slides cover a wide range of cellularity, but the data indicate that minimum adequate cellular counts for the SP and TP systems of 15,000 and 5000, respectively, appear appropriate in terms of excluding slides suitable for reading below these counts, as detection rates fell in samples below these levels of cellularity.

It can be reasonably concluded that a standardised cell counting protocol would be valuable, setting a minimum adequate cellular count at 15,000 for the SP system and 5000 for the TP system.

Scientific summary

Background

Liquid-based cytology (LBC) has been implemented across the UK Cervical Screening Programme following a National Institute for Health and Care Excellence (NICE) recommendation in 2003. There are two different cytology systems, SurePath™ (SP; BD Diagnostics, Burlington, NC, USA) and ThinPrep™ (TP; Hologic, Inc., Bedford, MA, USA), both of which are used in the UK. The NICE report highlighted the need for a definition of cell adequacy, following which a study designed to address this was commissioned by the National Institute for Health Research Health Technology Assessment programme. Previously reported studies on this topic have lacked sufficiently robust evidence on which to base the necessary practice guidelines. The Bethesda System guidance has recommended a minimum acceptable cell count (MACC) for the TP system of 5000 cells; however, this was reached in a rather arbitrary way and practice varies. As both systems are extensively employed in the NHS Cervical Screening Programme (NHSCSP), it was necessary that the required study evaluate both systems. This report describes the results of the study, which include a survey of current practice across 56 laboratories; investigations into the reliability of cell counting; the relationship between cell counts and slide reporting across these 56 laboratories; and the impact of serial dilution on slide reading. The aim of this study was to achieve a reliable basis on which recommendations could be developed for MACCs for both TP and SP.

Objectives

The objectives of this study were (1) to conduct a survey across the country of cytology laboratories in order to assess current standard procedures and practice for designating LBC slides inadequate; (2) to use a standardised counting method to achieve a reproducible basis of establishing the cellularity of a LBC sample; (3) to correlate the classification of reported slides (inadequate, negative or abnormal) among slides of known cellularity; and (4) to evaluate the impact of varying cellularity through dilution on the likelihood of detecting abnormal cells.

Methods

The overall study was divided into four separate components to address the objectives:

1. Current standard procedures and practice were determined through a questionnaire-based survey of 56 cytology laboratories in England, Wales and Scotland and through review of submitted existing protocols.
2. A counting protocol was developed by three experienced cytopathologists who each counted cells on a sample of routinely obtained slides to allow a comparison of counts using alternative starting points for cell counting in adjacent high-power fields. The total cell count was computed as (mean cell count of 10 high-power fields) × (area of cell deposit) ÷ (area of ocular). This allowed interobserver variation and the reproducibility of counts to be assessed, in order to select an optimal standard counting protocol.
3. A slide set was generated from all participating laboratories by requesting 20 slides from each of the inadequate, mild dyskaryosis and high-grade dyskaryosis classifications and a further 50 slides classed as negative. These were batched and recirculated to the network of 56 laboratories within the NHSCSP, Scottish Cervical Screening Programme and Cervical Screening Wales that had been recruited to survey. Cell counts were performed by designated staff in each laboratory using the standard counting protocol and counts were correlated with the reported slide classification (inadequate, negative, low or high grade).

4. Two methods of dilution were employed on cell samples containing dyskaryotic cells, in order to study the impact of reducing cellularity on the detection of cytological abnormalities. The first method was simple fluid dilution, termed unmixed dilution, and the second method involved mixing the sample with normal cells in order to retain overall cellularity but reduce the proportion of abnormal cells. The first method was intended to mimic overall hypocellular slides and the second method to mimic slides in which a diminishing number of abnormal cells were present within the entire range of cellularity.

Results

The survey of cervical screening laboratories included 28 SP and 28 TP laboratories and, of these, all of the SP laboratories and 27 TP laboratories responded. Practice was found to be variable; specifically, 15 out of 27 TP laboratories and 18 out of 28 SP laboratories used morphological criteria to determine slide adequacy, whereas 7 out of 27 TP laboratories and 13 out of 28 SP laboratories recorded indicators of transformation zone sampling, including both endocervical cells and metaplastic squamous cells. All but one of the SP laboratories assessed specimen adequacy by means of a MACC, and of the 11 out of 28 SP laboratories that specified a MACC figure, all stated a minimum of 15,000. Among the TP laboratories, 20 out of 27 assessed MACC by cell counting, and for the 11 out of 27 laboratories that specified the MACC figure, it ranged between 5000 and 15,000. Of the laboratories that responded to the survey, 29 out of 55 submitted standard operating procedures (SOPs).

The exercise to assess reliability of cell counting demonstrated a kappa value for SP of 0.851 [95% confidence interval (CI) 0.787 to 0.915] for counts performed in the same starting position and 0.906 (95% CI 0.804 to 1.00) in a different starting position. The corresponding TP figures were 0.614 (95% CI 0.461 to 0.767) and 0.590 (95% CI 0.407 to 0.774) for the same and different starting positions, respectively. There was, therefore, no significant difference between starting positions of counting but SP showed stronger interobserver agreement than TP. Numerically, there were very few instances of substantial disagreements in cell counts.

The cell counting survey from all participating laboratories amounted to 3110 SP slides and 3176 TP slides. The mean cell counts for inadequate samples were around 14,000 and 11,000 for SP and TP, respectively. A one-way analysis of variance (ANOVA) revealed a significant difference in these counts between SP laboratories but not between TP laboratories.

The cell counts for samples assessed as inadequate were far lower than the mean counts for negative, low- and high-grade abnormal samples, which averaged around 50,000. Of the SP slides submitted, 75% of inadequates had cell counts of less than 15,000, and only 2.5%, 2.3% and 1.4% of the high-grade, low-grade and negative samples, respectively, were below this count. A MACC set at 15,000 for SP would, therefore, achieve a sensible balance between sensitivity to detect cytological abnormalities and the maintenance of low rates for inadequate samples. With regard to TP, 43% of inadequate slides had cell counts of less than 5000 and only 1.8%, 1.6% and 1.3% of high-grade, low-grade, and negative samples, respectively, were below this count. Therefore, a MACC set at 5000 for TP would achieve a similar balance between sensitivity and inadequate rates.

The dilution study to vary slide cellularity involved 2400 slides of increasing dilution, which were each sent to three laboratories from a panel of 24. The overall kappa coefficient for slides of all categories was very similar for SP and TP at 0.593 (95% CI 0.571 to 0.610) and 0.609 (95% CI 0.589 to 0.633), respectively. For both systems the kappa value for the inadequate and the 'low- and high-grade'-combined samples was higher than for negative and high-grade samples. In the unmixed dilutions, there was evidence that for SP the detection of either low- or high-grade abnormalities increased as cellularity increased from 75% (95% CI 59.2% to 86.1%) at a cellularity of between 5000 and 10,000, to 91.7% (95% CI 87.4% to 94.6%), at a cellularity of over 50,000. For the unmixed dilutions in TP, there was a similar observation with cellularity below 2500 showing a detection of 75.8% (95% CI 55% to 88%) rising significantly to

98.2% (95% CI 91.5% to 99.7%) at cellularity between 50,000 and 75,000. The lowest 10% of cellularity (< 15,000 for SP and < 5000 for TP) had a significantly lower detection of abnormalities in both systems than the middle 80% (up to 58,000 for both SP and TP).

In the mixed-dilution study, it was shown that, as the number of dyskaryotic cells was diluted by the number of normal cells, the likelihood of detecting abnormalities was reduced significantly in SP but not significantly in TP. As the actual number of dyskaryotic cells decreased so did the likelihood of detection. Compared with a reference standard of greater than 50 dyskaryotic cells on slides above the MACC, the odds ratio for detection when fewer than 25 dyskaryotic cells were present was 0.49 and 0.74 for SP and TP, respectively.

Conclusions

Each section of this study has yielded the basis for reliable conclusions which have implications for laboratory practice within the NHSCSP. There is quite variable practice in classifying slides as inadequate, and it would be beneficial to agree a standard procedure for cell counting, and a standard MACC for both SP and TP. The counting SOP specified in this report showed good reproducibility and could be widely adopted. It was clear that there is a very wide range of cell counts for slides classified as inadequate, some of which are a result of factors other than hypocellularity. The evidence from this study suggests that a MACC of 15,000 and 5000 for SP and TP, respectively, would probably achieve the best balance between maintaining low levels of inadequate slides and not compromising on the chances of detecting abnormalities. Many inadequate slides had counts above this recommended MACC. However, a MACC threshold above which there would be few inadequate samples would risk a significant rise in the proportion of inadequate slides reported, without any evidence of improved detection of abnormal cells. Finally, there is evidence that in slides with a normal range of cell counts, slides with fewer than 25 dyskaryotic cells are associated with a reduced chance of detecting abnormalities compared with slides with more than 50 dyskaryotic cells, and this may have implications for medicolegal cases where missed abnormalities have resulted in the development of cancer.

Future work

The findings of this study should inform the development of laboratory practice guidelines.

Funding

The National Institute for Health Research Health Technology Assessment programme.

Chapter 1 Introduction

Exfoliative cervical cytology has formed the basis of cervical screening since the 1960s. It is universally accepted that, where systematic population-based screening programmes have been established, the incidence and mortality rate from cancer of the cervix have fallen as a direct consequence. Examples of this include British Columbia,[1] England[2] and Denmark.[3] The rationale of cervical screening is that regular screening every 3–5 years by means of cytology and subsequent colposcopy can lead to a diagnosis of pre-malignant lesions known as high-grade cervical intraepithelial neoplasia (CIN), treatment of which prevents cancer. The presence of CIN grade 3 is regarded as the true precursor lesion of squamous cell carcinoma of the cervix, but CIN grade 2 is the commonly used threshold for treatment. It is generally agreed that the sensitivity of a single conventional cytology test to detect underlying CIN is around 50–70%;[4] therefore, repeated cytology is required at regular intervals and this is considered to prevent around 70% of cervical cancers.[5] Until recently, the standard method used for cervical cytology was termed a cervical smear because cells were scraped from the cervix using a wooden or plastic spatula or a brush sampling device and spread or smeared onto a glass slide, fixed in alcohol, stained as described by Papanicolaou[6] and viewed under a microscope. This traditional technique remained unchallenged for almost 50 years, but there were problems with blood, inflammatory cells and debris obscuring the epithelial cells. In addition, clumping of epithelial cells, as well as scanty cellularity due to poor transfer of cellular material from the sampling device to the glass slide, could result in the sample being classified as 'unsatisfactory' or 'inadequate', requiring a repeat sample to be taken. Compared with international practice, inadequate rates were a particular problem in the UK, possibly because of rigorous reporting standards, requiring 7–8% of women to reattend for screening, which was inconvenient and wasteful.

During the 1990s, a new technology was developed, known as liquid-based cytology (LBC). This relied on the cervical sample being obtained by a brush sampling device, which was then rinsed or placed in an alcohol-based liquid transport medium. An aliquot of the homogenised liquid sample was then used to produce a cellular deposit on a glass slide. Two technologies, namely ThinPrep™ (TP; Hologic, Inc., Bedford, MA, USA) and SurePath™ (SP; BD Diagnostics, Burlington, NC, USA), currently dominate the market. Although each technology produces cleaner more homogeneous preparations, with less obscuring of diagnostic material, and hence lower inadequate rates,[7] the methodologies by which this is achieved are quite distinct. The TP system aspirates the liquid sample through a filter until a programmed quantity of cellular material has been acquired and this is deposited on a glass slide. The slide is then stained on separate technology prior to screening. The SP system uses a sequential process of sample enrichment and sedimentation to produce a cellular deposit, which is individually stained on the LBC platform.

Around the time LBC was ready for clinical use, it was becoming clear that human papillomavirus (HPV) testing would play a major role in cervical screening because of increased sensitivity and the opportunity for extended screening intervals. In addition, it could exploit the high negative predictive value of HPV, in order to streamline protocols such as triage of low-grade abnormalities and test of cure following treatment of CIN. This acted as a spur to evaluate LBC, which would enable HPV testing to be reflexly triaged by cytology, or vice versa, without having to obtain a second sample. NHS pilot studies of LBC were reported in 2003,[7] which demonstrated a major reduction in the rates of inadequate slides, and an economic evaluation confirmed LBC to be cost-effective even though it cost more than conventional cytology. A pooled analysis, based on seven trials, published in 2008 concluded that LBC was neither more sensitive nor more specific in terms of detection of high-grade CIN than conventional cytology.[8]

The National Institute for Health and Care Excellence (NICE) considered LBC in 2003[9] and recommended its national implementation, which was completed by 2008. The NICE report highlighted the need to determine if there was a threshold of cellularity, which should be determined to define the adequacy of a slide.

In the USA, the Bethesda System (TBS) required a minimum of 5000 squamous cells on the slide for a preparation to be regarded as adequate.[10] Since that time a number of studies have addressed this issue, but there have been two problems in defining cell adequacy. The first is the absence of a reliable and widely used method of cell counting. The second is the lack of robustly designed prospective studies replicating real-life practice, which could provide a reliable evidence base for a broadly acceptable definition of cell adequacy in LBC.

One rather elegant study using TP LBC[11] established that 87 abnormal cells on a slide were required to achieve 98% sensitivity, in terms of detecting severe dyskaryosis. In a slide with 5000 squamous cells, the authors, therefore, surmised that the ratio of of abnormal to normal cells would be1 : 47. In reading slides in which abnormalities were detected, this ratio rate ranged from 1 : 2.5 to 1 : 4596. These authors reasonably concluded that it is unlikely that a precise cellularity threshold could be established which achieved both minimal risk of missing an abnormality and achievement of a minimum number of rejected slides because of hypocellularity. They went on to conclude that 5000 cells in TP could achieve an inadequate rate of less than 5% as well as sufficient sensitivity. Another study,[12] which was reported only as a conference abstract, used dilutions of SP specimens to determine if there was a valid threshold in cellularity in terms of maintaining high sensitivity of SP. Such a demarcation was reported to exist at around 5000 cells.[12] This contrasts sharply with a recently reported study using the TP system from the Netherlands,[13] in which, based on seven assessments for adequacy, a majority score of 'unsatisfactory' or 'satisfactory but limited by scant cellularity' was found in 42 cases, 41 of which had a cell count of less than 20,000. In this study, the most accurate cell counting protocol was found to be based on counting five non-adjacent microscope fields along a horizontal axis, and five along the vertical axis using a ×10 objective and applying a correction factor of 1.14 for underestimation of the true cellularity.

It is, therefore, apparent that there is currently no universal agreement on either a cell adequacy threshold or a reliable protocol for counting cells on a slide. The present study was performed to assess the variation in assessment for cellular inadequacy, and also to try to establish a reliable threshold for defining inadequacy based on cell count which could be applied across the NHS Cervical Screening Programme (NHSCSP).

Objectives stated in study protocol

1. To assess current standards and practice for the reporting of LBC preparations across England, Scotland and Wales.
2. To establish a reproducible method for rapidly estimating the cellularity of a LBC sample.
3. To determine the cellularity of samples classified inadequate, negative or abnormal by a range of laboratories across the country.

    - To assess the impact of varying the overall cellularity on the likelihood of detection of cytological abnormalities
    - To assess the impact of varying the relative proportion of abnormal cells on the likelihood of detection of cytological abnormalities.

Objective 1: to assess current standards and practice for the reporting of liquid-based cytology preparations across England, Scotland and Wales

Survey of working practice

In November 2007, questionnaire surveys (see *Appendix 1*) to assess the standards and practice of reporting LBC adequacy were sent to the 56 laboratories in England, Scotland and Wales that had agreed to participate in the study. In Scotland only TP was used, in Wales only SP was used and in England both SP and TP were used. All but one (98%) of the questionnaires were returned. Of those that responded, 28 used SP (*Table 1*) and 27 laboratories used TP (*Table 2*). Of these laboratories, 15 out of 28 (54%) of the SP laboratories and 14 out of 27 (52%) of the TP laboratories also provided a copy of their standard operating procedure (SOP) for assessment of specimen adequacy.

TABLE 1 Standards and practice of LBC reporting: SP laboratories in England and Wales – 2007

Laboratory[a]	Morphological adequacy criteria	Transformation zone criteria	Cell counting methodology	Minimum number of squamous cells
A	No	Yes	Minimum number squamous cells per ocular field	15,000
B	Yes	No	Minimum number squamous cells per ocular field	15,000
C	Yes	Yes	Minimum number squamous cells per ocular field	15,000
D	No	No	Minimum number squamous cells per ocular field	15,000
E	No	No	Minimum number squamous cells per ocular field	15,000
F	No	No	Minimum number squamous cells per ocular field	15,000
G	Yes	No	Minimum number squamous cells per ocular field	Not stated
H	Yes	No	Not provided	Not stated
I	Yes	No	Minimum number squamous cells per ocular field	Not stated
J	Yes	Yes	Minimum number squamous cells per ocular field	15,000
K	Yes	Yes	Minimum number squamous cells per ocular field	Not stated
L	Yes	Yes	Minimum number squamous cells per ocular field	Not stated
M	No	No	Minimum number squamous cells per ocular field	Not stated
N	No	No	Minimum number squamous cells per ocular field	15,000
O	Yes	No	Minimum number squamous cells per ocular field	Not stated
P	No	Yes	Minimum number squamous cells per ocular field	Not stated

continued

TABLE 1 Standards and practice of LBC reporting: SP laboratories in England and Wales – 2007 (*continued*)

Laboratory[a]	Morphological adequacy criteria	Transformation zone criteria	Cell counting methodology	Minimum number of squamous cells
Q	No	Yes	Minimum number squamous cells per ocular field	Not stated
R	No	Yes	Minimum number squamous cells per ocular field	15,000
S	Yes	No	Minimum number squamous cells per ocular field	Not stated
T	No	No	Minimum number squamous cells per ocular field	Not stated
U	Yes	Yes	Minimum number squamous cells per ocular field	Not stated
V	Yes	No	Minimum number squamous cells per ocular field	Not stated
W	Yes	Yes	Minimum number squamous cells per ocular field	Not stated
X	Yes	No	Minimum number squamous cells per ocular field	15,000
Z	Yes	Yes	Minimum number squamous cells per ocular field	Not stated
AA	Yes	No	Minimum number squamous cells per ocular field	Not stated
BB	Yes	Yes	Minimum number squamous cells per ocular field	15,000

a Scotland has never used the SP system and LBC had not been introduced in Northern Ireland at this time.

TABLE 2 Standards and practice of LBC reporting: TP laboratories in England and Scotland – 2007

Laboratory[a]	Morphological adequacy criteria	Transformation zone criteria	Cell counting methodology	Minimum number of squamous cells
A	No	No	Not provided	Not stated
B	Yes	Yes	Minimum number squamous cells per ocular field	12,000
C	No	No	Minimum number squamous cells per ocular field	10,000
D	Yes	Yes	Not provided	Not stated
E	Yes	No	Minimum number squamous cells per ocular field	8000–10,000
F	Yes	No	Minimum number squamous cells per ocular field	Not stated
G	Yes	No	Minimum number squamous cells per ocular field	Not stated
H	Yes	No	Minimum number squamous cells per ocular field	8000–10,000
I	Yes	No	Minimum number squamous cells per ocular field	Not stated

TABLE 2 Standards and practice of LBC reporting: TP laboratories in England and Scotland – 2007 (*continued*)

Laboratory[a]	Morphological adequacy criteria	Transformation zone criteria	Cell counting methodology	Minimum number of squamous cells
J	Yes	Yes	Minimum number squamous cells per ocular field	5000
K	Yes	No	Not provided	5000
L	Yes	No	Minimum number squamous cells per ocular field	Not stated
M	Yes	No	Minimum number squamous cells per ocular field	13,000
N	No	No	Minimum number squamous cells per ocular field	15,000
O	No	No	Minimum number squamous cells per ocular field	Not stated
P	No	No	Not provided	Not stated
Q	Yes	Yes	Minimum number squamous cells per ocular field	Not stated
R	No	Yes	Minimum number squamous cells per ocular field	Not stated
S	No	No	Not provided	Not stated
T	No	No	Minimum number squamous cells per ocular field	15,000
U	No	Yes	Minimum number squamous cells per ocular field	Not stated
V	No	No	Minimum number squamous cells per ocular field	5000
W	Yes	No	Minimum number squamous cells per ocular field	Not stated
X	No	No	Not provided	Not stated
Y	Yes	Yes	Not provided	Not stated
Z	Yes	No	Minimum number squamous cells per ocular field	Not stated
AA	No	No	Minimum number squamous cells per ocular field	9000–11,000

a Wales did not use TP at this time.

Among the SP laboratories, 18 out of 28 (64%) stated that they use morphological criteria (epithelial cells obscured by mucus or blood; cytolysis; absence of endocervical cells in follow-up of cervical glandular intraepithelial neoplasia) to determine specimen adequacy and 13 out of 28 (46.4%) stated that they record the presence of indicators of transformation zone sampling (endocervical cells or metaplastic squamous epithelial cells). The vast majority (27 out of 28, 96.4%) assess specimen adequacy by counting a minimum number of squamous epithelial cells in adjacent ocular fields. Among those in the SP laboratories that stated the calculated/estimated minimum number of squamous epithelial cells that must be present on a slide for it to be regarded as adequate (11 out of 28), all gave a figure of at least 15,000 cells.

Of the TP laboratories surveyed, 15 out of 27 (56%) stated that they use morphological criteria (epithelial cells obscured by mucus or blood; cytolysis; absence of endocervical cells in follow-up of cervical glandular intraepithelial neoplasia) or technical criteria (sampling brush left in specimen container) to determine specimen adequacy. The presence of indicators of transformation zone sampling (endocervical cells or metaplastic squamous epithelial cells) was recorded by 7 out of 27 (25.9%) laboratories. A majority of the laboratories (20 out of 27, 74%) assess specimen adequacy by counting a minimum number of squamous epithelial cells in adjacent ocular fields. Among the TP laboratories, for the 11 out of 27 (41%) that stated the calculated/estimated minimum number of squamous epithelial cells that must be present on a slide for it to be regarded as adequate, the range of 5000–15,000 cells was counted, usually in response to local or regional guidance. The distribution of the minimum calculated/estimated number of squamous epithelial cells was as shown in *Table 3*.

Discussion

In summary, while the majority of laboratories using TP and SP LBC systems use morphological and, in the case of TP, technical criteria to assess specimen adequacy, most do not record the presence of indicators of transformation zone sampling. The majority of laboratories assess specimen adequacy by counting or estimating a minimum number of squamous epithelial cells. All SP laboratories require a minimum of 15,000 cells for a sample to be considered adequate, but in TP laboratories there is a wide range of minimum acceptable cellular counts (MACCs), varying from 5000 to 15,000 cells, with only 3 out of 11 laboratories using a MACC of 5000.

This variation in practice is strong justification for a study to determine a MACC for both SP and TP LBC systems and provide guidance for the NHSCSP.

The background to this study was the NICE report that highlighted the need to determine whether or not there was a threshold of cellularity which should be determined to define the adequacy of a slide.[9] In the interim, UK laboratories had adopted a figure of up to 15,000 cells as a determinant of adequate LBC cellularity for the SP system, based on the LBC pilot experience and other guidance, and this was confirmed in the questionnaire survey. For all laboratories, an estimation of total cellularity was obtained by performing representative ocular field cell counts, although a review of the submitted SOPs revealed that practice varied from one laboratory to another; that is, truly adjacent consecutive fields or fields with spaces between starting point and direction of counting. Specific guidance for Scotland, where the TP samples system is exclusively used, recommended a minimum of 10,000 well-preserved squamous epithelial cells, and this was confirmed in the survey. All SP laboratories used a minimum of 15,000 cells.

TABLE 3 Distribution of minimum cellular criteria for TP among respondent laboratories

Minimum number of squamous cells	Number of laboratories
5000	3
8000–11,000	2
9000–11,000	1
10,000	1
12,000	1
13,000	1
15,000	2
Total	11

In addition, a majority of both SP and TP laboratories used morphological indicators to assess specimen adequacy. A substantial minority of both SP and TP laboratories recorded the presence of indicators of transformation zone sampling, but it would appear that this criterion was not used in the majority of laboratories as an indicator of an adequate sample.

More recently, pending the results of this study, quality assurance guidance for the NHSCSP has recommended that an adequate liquid-based sample is defined as one that contains the minimum level of squamous epithelial cellularity necessary to ensure a squamous abnormality detection rate equivalent to that offered by conventional smears.[14]

It is clear from this survey that differences in practice exist, that these differences are not evidence based, and that it is timely to develop evidence-based practice guidelines. Developing the evidence required to produce such guidance was the purpose of this study, as described in *Objectives 2*, *3* and *4*, which in turn cover counting methodology, cell counts from all of the study laboratories on a standard slide set and the use of cell dilution to evaluate the effect of reducing overall cell counts, as well as dyskaryotic cell counts, on the chances of detection.

Objective 2: development and test of cell counting methodology

Method

The objective was to establish a reliable method for rapidly estimating the cellularity of a LBC sample for SP and TP slides. Cells were included in the count if they were intact, mature or parabasal squamous cells with nuclei, even if the nuclei were pale. Syncytial aggregates of squamous cells, as seen in cytolysis, were counted according to the number of nuclei they contained, even if the cytoplasmic margins of individual cells were not identifiable. Any free nuclei, and any anucleate squamous cells or fragments of squamous cytoplasm were not counted. If no cellular material was present then a zero value was recorded. If exceptionally thick groups of cells were present, an estimate of cellularity was used. A full quadrant of a high-power microscope field (×40 objective) contains approximately 1000 small parabasal squamous cells and 750 mature squamous cells. Cells at the edge of the field were counted if the entire circumference of the nucleus was seen (otherwise they were not counted).

The cell count for a slide was obtained by counting the number of squamous cells in 10 fields of view. The total cell count for a slide was then computed as

$$\text{Total cell count} = \frac{(\text{mean cell count of 10 fields of view}) \times (\text{area of cell deposit})}{(\text{area of ocular})} \quad (1)$$

The SOP by which the 10 fields of view were selected was as follows. An assessor had a choice of one of four starting points to begin the slide count. Assuming that the deposit represented a clock face, a starting position on the perimeter at 12, 3, 6, or 9 o'clock was chosen for the first field that was neither hypo- nor hypercellular. After counting cells in the starting field of view, the assessors then moved along a radius towards the centre of the slide in steps determined by the width of the field of view counting a further nine fields. For the SP LBC system, adjacent fields of view were counted, whereas for the TP system every second field was counted (see *Appendices 3* and *4*).

Reliability study of squamous cell counting procedure

A reliability study was carried out to assess the consistency of the cell counting procedure. This considered two situations. The first was a pair of assessors using the same starting position (12, 3, 6, or 9 o'clock), and the second was a pair of assessors using different starting positions. For each LBC system, 30 slides were assessed by three experienced cytopathologists (LT, MD and JS). LT performed the cell count procedure for all four starting points, completing 120 cell counts for each system. So as to mimic the everyday laboratory procedure, JS and MD performed just one cell count for each slide using their choice of starting position based on the cell counting SOP, thereby completing 30 cell counts each for each system. This gave data that enabled the repeatability of the procedure to be assessed when using the same starting point or different starting points.

Statistical analysis (total cell counting)

Reliability of cell counting was assessed using an intraclass correlation coefficient (ICC). In this setting, the ICC estimates the proportion of the total variance of the cell count between slides. In the ideal situation, where there is no measurement error, that is no disagreement between assessors/cytopathologists, the ICC will be equal to 1. Alternatively, if all the measurement was noise, so that nothing is being learnt regarding the true cellularity, the ICC would be equal to 0. An ICC closer to 1 therefore represents higher reliability. To calculate the ICCs provided by three assessors using same/different starting positions, nine pairs of results (four pairs JS with LT, four pairs MD with LT, and one pair JS and MD) were created for each slide and corresponding confidence intervals (CIs) calculated using a bootstrap procedure. A bootstrap procedure is performed by randomly sampling with replacement of each pair of results until a sample of equivalent size to the original data is created. This is repeated 1000 times. Each sample is then analysed and the results are combined together to give robust standard errors.[15] For each LBC system, the total cellularity of 30 slides was assessed, with LT completing 120 cell counts and JS and MD each completing 30. JS and MD used the same starting point for 8 of the 30 SP slides and 15 of the 30 TP slides.

Statistical analysis (thresholds of cellularity)

We began by comparing the agreement of assessors in determining if the cellularity was above or below the MACC. Agreement was measured using a kappa coefficient, and the same pairing and bootstrapping procedure as for the previous analysis using the ICC. The kappa coefficient is a measure of reliability for categorical data corresponding to the ICC used for the total cell count and has a similar interpretation. A kappa coefficient of 1 implies complete agreement between the two ratings, so there is no measurement error, whereas a kappa coefficient of 0 implies no agreement beyond that as a result of chance.

Results

Reliability of total cell counts

The ICCs for all SP slide assessments were 0.547 (95% CI 0.456 to 0.638) for both same and different starting positions, 0.712 (95% CI 0.603 to 0.821) for those performed in the same starting position and 0.505 (95% CI 0.404 to 0.605) for those performed in different starting positions. The corresponding TP ICCs were 0.741 (95% CI 0.700 to 0.784) for both the same and different starting positions, 0.750 (95% CI 0.651 to 0.850) for same starting position and 0.740 (95% CI 0.692 to 0.788) for observations in different starting positions.

For SP and TP, reliability was, therefore, higher where the two assessors used the same starting position. While this is to be expected, it should be noted that there is some overlap of the CIs.

Reliability of counting at thresholds of cellularity

The analysis above considered the reliability of a total cell count. Perhaps more important in clinical practice is the reliability of cell counting close to the values that might be used to consider adequacy. For the SP system, we categorised the cell count as < 10,000, 10,000–14,999 and ≥ 15,000, and for TP we used the banding < 5000, 5000–10,000 and ≥ 10,000. *Table 4* summarises the pairs of assessors scores. There were 270 ratings in total, resulting from 30 slides in nine pair comparisons (4 = LT and JS, 4 = LT and MD and 1 = JS and MD). Of the 270 pairs of assessments using the SP system, both cytopathologists classified the slide as having a cell count above 15,000 in 94 pairs. *Table 4* shows that, for SP, assessors disagreed over 19 slide assessments with one assessment being above 15,000 and the other below. Similarly, for TP, there was disagreement over eight slide assessments with one scoring above 10,000 and the other less than 5000.

The kappas associated with the MACC cellularity threshold for all SP slide assessments were 0.851 (95% CI 0.787 to 0.915) for both the same and different starting positions, 0.906 (0.804 to 1.00) for those performed in the same starting position and 0.832 (95% CI 0.752 to 0.913) for those performed in a different starting position. The corresponding TP kappas at a MACC of 5000 were 0.614 (95% CI 0.461 to 0.767) for both same and different starting positions, 0.685 (95% CI 0.350 to 1.00) for same starting position and 0.590 (95% CI 0.407 to 0.774) for observations in different starting positions, and at a MACC of 10,000 were 0.657 (95% CI 0.549 to 0.766), 0.605 (95% CI 0.394 to 0.816) and 0.678 (95% CI 0.552 to 0.805), respectively.

TABLE 4 Comparison of current minimum adequate cellularity designation in pairs of raters when raters use same starting position, different starting positions and both the same and different starting positions

SP					TP				
Cellularity cut-off point	< 10,000	10,000–14,999	≥ 15,000	Total	Cellularity cut-off point	< 5000	5000–9999	≥ 10,000	Total
All assessments compared									
< 10,000	107	13	**3**	123	< 5000	19	5	**4**	28
10,000–14,999	9	28	**6**	43	5000–9999	7	15	18	40
≥ 15,000	**0**	**10**	94	104	≥ 10,000	**4**	7	191	202
Total	116	51	103	270[a]	Total	30	27	213	270[a]
Same starting position assessments only									
< 10,000	27	4	0	31	< 5000	5	1	1	7
10,000–14,999	3	7	1	11	5000–9999	0	7	5	12
≥ 15,000	0	2	24	26	≥ 10,000	2	3	51	56
Total	30	13	25	68	Total	7	11	57	75
Different starting position assessments only									
< 10,000	80	9	3	92	< 5000	14	4	3	21
10,000–14,999	6	21	5	32	5000–9999	7	8	13	28
≥ 15,000	0	8	70	78	≥ 10,000	2	4	140	146
Total	86	38	78	202	Total	23	16	156	195

a 270 ratings result from 30 slides in nine pair comparisons (4 = LT and JS, 4 = LT and MD and 1 = JS and MD).
The numbers in bold represent slides with disagreement between raters above and below 15,000 for SP and 10,000 for TP.

Discussion

The estimates of the kappa coefficient for cell counting are systematically lower for TP than for SP, but it should also be noted that the CIs for TP are wide and some overlap the CIs for SP. This difference between the systems may be explained by the smaller proportion of slides below the threshold used for TP. Low prevalence of slides below the threshold may explain the lower values of kappa, as binary scales with a prevalence closer to either 0 or 1 tend to have smaller values of the kappa coefficient than scales with prevalence closer to 0.5.[16] This property of the kappa coefficient has implications for the values of kappa observed for the SP threshold. In service settings, the prevalence of slides with cellularity below the threshold needs to be low if the screening method is to have utility. If the reliability study for SP were repeated in a sample with a more realistic proportion of subjects being below the threshold, it is likely that the kappa coefficient would be rather lower. Hence, the low value of kappa observed for TP may be more realistic than the value of kappa observed for SP. This has implications for the application of a strict threshold to this method of cell counting.

There is some evidence from *Table 4* of only a small proportion of substantial disagreements between assessors. Considering first SP, the number of occasions for which the two assessors disagreed substantially, that is, where one assessment was less than 10,000 and the other was greater than 15,000, was only three (see *Table 4*). Similarly, for TP, there were only eight (see *Table 4*) pairs of assessments where one assessor scored less than 5000 and the other above 10,000.

This study has some limitations. First, the sample size used for either system is small, with each being based on only 30 slides. A study with a larger sample size would be needed to gain a more precise estimate of reliability. A second issue is the representativeness of the slides used for the reliability exercise. The proportion of assessments with a cell count below 15,000 cells for SP or below 5000 cells for TP was higher than would be found in a representative service sample of all slides. Nonetheless, it should be noted that cell counting is likely to be applied as a formal technique only in samples that have been pre-screened 'informally'. The sample may, therefore, be rather more representative of samples to which one might actually apply formal cell counting. One avenue for further work would, therefore, be to compare the reliability of a combined formal and 'informal' assessment of slide cellularity.

A methodology for estimating the total number of epithelial cells on a slide by counting the number of epithelial cells within 10 fields of view and combining them to give an estimate of the total cell count has been developed and evaluated in this study. To assess the reproducibility of this method, a reliability study was carried out which revealed an ICC between 0.547 (95% CI –0.456 to 0.638), representing moderate agreement in evaluation of TP samples, and 0.741 (95% CI 0.700 to 0.784), representing strong agreement in evaluation of SP samples. This difference might be expected because of the more uniform distribution of cells on SP slides, and it may also explain why in TP samples the point estimates of reliability in the comparisons of the same starting position are less than in different starting positions. For the same reason, it is surprising that the ICCs for the two assessors who followed the SOP strictly regardless of the choice of starting position for counting were 0.305 (95% CI 0.000 to 0.633 – fair agreement) and 0.650 (95% CI 0.441 to 0.858 – moderate agreement) for SP and TP, respectively. It should also be noted that the cell counts were performed by three senior experienced consultant cytopathologists, whereas in routine practice it is expected that these counts would be performed by primary screening staff, either cytology screeners or biomedical scientists, and these staff groups have different microscopic and morphological interpretive skill sets. It might have been better if this part of the study had been undertaken by primary screening staff in order to provide assurance that the counting methodology was appropriate and utilisable in routine clinical practice. However, in routine practice, formal cell counting as described is rarely required in SP samples, as the vast majority of samples are clearly adequate in terms of a naked-eye assessment, when the colour and density of the cell deposit strongly contrasts with inadequate or potentially inadequate samples. Inadequate samples show very pale-stained or virtually invisible cell deposits, and this is confirmed on low-power microscopic examination.

In terms of estimating LBC cellularity, there appears to be a better interassessor agreement with SP than with TP, and this may be related to the qualitative differences between the way the cells are presented on the slide. There also appear to be differences in reproducibility in terms of counting; however, it is not possible to state categorically that a specific method (same or different starting position) achieves greater reliability. Given the variation inherent in cell counting, it would perhaps be optimal if an agreed counting method were agreed for use in the NHSCSP.

Objective 3: a survey of slide cellularity

Method

A survey was carried out to determine the distribution of cellularity of samples classified as inadequate, negative or abnormal. The objectives of the survey were:

- to determine the distribution of cellularity of samples classified as inadequate, negative or abnormal
- to investigate the threshold used by different laboratories to determine adequate cellularity
- to investigate the cellularity distribution of samples classified as cytology negative and HPV positive.

The 56 laboratories which were involved in the survey (28 SP and 28 TP) were asked to submit 20 consecutive cervical screening LBC cases from each of the categories of inadequate, mild dyskaryosis and high-grade dyskaryosis (moderate dyskaryosis and above), and a further 50 consecutive negative cases. Some laboratories varied slightly from the requested numbers.

It was not feasible for all material from the slide survey to be assessed for cellularity at a single centre. The task of cell counting was, therefore, distributed across all participating laboratories. In order to reduce potential bias as a result of interlaboratory variability in the assessment of cellularity, each laboratory was sent a set of slides with similar composition by type (inadequate, mild dyskaryosis, high-grade dyskaryosis or negative) from the other laboratories using the same LBC system. Slides from the originating laboratories were relabelled with an anonymous code and repackaged, randomly sorted within the laboratory and by slide type before being systematically allocated to laboratories using a computer-generated pre-prepared list and then sent to the participating laboratory for cell counting. Each laboratory received approximately four slides from every other laboratory using the same system, with approximately 20 slides from each of the inadequate, mild dyskaryosis and high-grade dyskaryosis classifications and 50 slides classed as negative. Each laboratory was asked to nominate primary screening staff to carry out the cell counting. One laboratory withdrew from the study at this stage and their batch of slides was randomly reassigned to other laboratories in the study. All slides in the study sets were assessed for the presence of transformation zone indicators and each had a formal cell count according to the study cell counting SOP (see *Appendices 3* and *4*; see also *Objective 2: development and test of cell counting methodology*). Cell counts were recorded directly into a database to minimise data errors and the database was returned on completion to the study centre (Liverpool) with the slide sets.

Slides classified as cytology negative but HPV positive might be considered to contain a small proportion of cytological false negatives. Furthermore, if HPV primary screening is introduced, it is likely to require women who are HPV positive to undergo reflex cytology in order to triage ongoing management. To investigate the cellularity distribution of such samples classified as cytology negative but HPV positive, 1200 cases from the ARTISTIC (A Randomised Trial In Screening To Improve Cytology) study[17] previously documented as cytology negative and HPV positive were also subject to cell counting. A comparison of the cell count distribution with both the negative and mildly dyskaryotic cases from the slide survey would help to determine whether or not their potentially false-negative cytology relates to their cellularity. The ARTISTIC study used the TP LBC system; therefore, these cell counts were compared with those for negative and mildly dyskaryotic slides from the 28 TP laboratories.

Statistical analysis (slide comparisons)

The SP and TP total squamous cell counts across all slides and within each slide diagnosis (inadequate, high grade, mild dyskaryosis and negative) were compared using an independent-samples *t*-test. Owing to the highly skewed nature of the cell counts, a square root transformation was performed prior to the comparison tests. In order to describe the cell count distribution, continuous cell counts for both LBC methods were categorised as was deemed suitable by the cell count distribution and displayed in a cross-tabulation with slide diagnosis. For comparison, all TP tables also include cell counts from the cytopathology-negative/HPV-positive ARTISTIC data set.[17]

Assessing consistency across reading laboratories within inadequate and negative slides is important, and comparisons of cellularity for inadequate slides across reading laboratories were performed. A one-way analysis of variance (ANOVA) compared mean cellularity between laboratories for inadequate slides and a chi-squared test compared the proportion of slides with cellularity above the cut-off points 15,000 and 5000 between laboratories. Formal statistical testing is difficult to interpret when comparing large numbers of units because of the large number of post-hoc pairwise comparisons that can be made. The ICC was performed as an alternative measure of heterogeneity in order to assess the magnitude of variation between laboratories in either the mean cellularity of inadequate slides or the proportion of inadequate slides above the MACC threshold when compared with the total variation. The ICC estimates the proportion of the total variation that occurs between laboratories. In this context, an ICC of 0 indicates that there is no variation between laboratories above that as a result of sampling variation. Small sample sizes when comparing the cellularity of negative slides by laboratory means that the expected cell counts are likely to be small, hence a Fisher's exact test may replace the chi-squared test.

Results

Almost all laboratories submitted the minimum number of requested slides (110 slides). Two TP laboratories produced 123 slides and three SP laboratories produced 123, 125 and 136 slides. In all, 3110 cell counts were carried out on SP slides and 3176 cell counts on TP slides, totalling 6286 slides.

Distribution of cellularity by slide type

Table 5 gives the mean [standard deviation (SD)] cellularity score and range for inadequate, mild dyskaryosis, high-grade dyskaryosis or negative slides for both SP and TP. Cell counts were statistically significantly higher for SP than for TP for all four categories. *Table 6* gives the frequency and cumulative frequency distribution of cell counts by slide type for SP. Of the inadequate slides, 74.5% had a cell count of less than 15,000, whereas only 1.9% (47 out of 2522) of adequate slides reported as negative or low/high grade were below that level. More than 90% of dyskaryotic or negative slides had a cell count above 25,000.

TABLE 5 Summary statistics of cellularity and test comparison between LBC system

Slide diagnosis	SP				TP				*t*-test *p*-value[a]
	n	Mean	SD	Range	*n*	Mean	SD	Range	
Inadequate	587	13,979.7	15,586.1	0–152,181	623	10,870	14,824.7	0–140,789	< 0.001
High-grade dyskaryosis	559	54,419.2	25,129.3	2072–177,880	572	43,323.9	27,033.7	0–153,781	< 0.001
Low-grade dyskaryosis	561	53,054.2	23,332	542–190,590	562	51,120.4	31,822.2	398–234,119	0.009
Negative	1402	58,163.2	32,748	2352–313,045	1418	49,284.5	33,812.7	265–332,353	< 0.001
Total	3109	48,226	31,903.7	0–313,045	3175	40,997.9	33,043.9	0–332,353	< 0.001
Cytological negative/HPV positive	–	–	–	–	1200	52,166.8	34,453.3	928–259,174	–

a *t*-test based on square root-transformed cellularity because of positive skewness.

TABLE 6 Distribution of cellularity by original slide diagnosis (SP)

	Inadequate		High-grade dyskaryosis		Low-grade dyskaryosis		Negative	
Cellularity	Frequency (%)	Cumulative%	Frequency (%)	Cumulative %	Frequency (%)	Cumulative %	Frequency (%)	Cumulative %
0–2499	44 (7.5)	7.5	1 (0.2)	0.2	1 (0.2)	0.2	1 (0.1)	0.1
2500–4999	60 (10.2)	17.7	0 (0)	0.2	1 (0.2)	0.4	1 (0.1)	0.1
5000–7499	80 (13.6)	31.4	1 (0.2)	0.4	0 (0)	0.4	0 (0)	0.1
7500–9999	90 (15.3)	46.7	2 (0.4)	0.7	2 (0.4)	0.7	3 (0.2)	0.4
10,000–14,999	163 (27.8)	74.5	10 (1.8)	2.5	9 (1.6)	2.3	15 (1.1)	1.4
15,000–19,999	57 (9.7)	84.2	5 (0.9)	3.4	16 (2.9)	5.2	29 (2.1)	3.5
20,000–24,999	26 (4.4)	88.6	27 (4.8)	8.2	16 (2.9)	8	45 (3.2)	6.7
25,000–29,999	18 (3.1)	91.7	31 (5.5)	13.8	27 (4.8)	12.8	74 (5.3)	12
30,000–39,999	23 (3.9)	95.6	90 (16.1)	29.9	95 (16.9)	29.8	227 (16.2)	28.2
40,000–49,999	6 (1)	96.6	103 (18.4)	48.3	116 (20.7)	50.4	267 (19)	47.2
50,000–74,999	13 (2.2)	98.8	190 (34)	82.3	198 (35.3)	85.7	466 (33.2)	80.5
75,000–99,999	3 (0.5)	99.3	68 (12.2)	94.5	62 (11.1)	96.8	163 (11.6)	92.1
100,000–149,999	3 (0.5)	100	27 (4.8)	99.3	17 (3)	100	80 (5.7)	97.8
150,000–199,999	1 (0.2)	100	4 (0.7)	100	1 (0.2)	100	19 (1.4)	99.2
200,000–249,999	0 (0)	100	0 (0)	100	0 (0)	100	8 (0.6)	100
250,000–299,999	0 (0)	100	0 (0)	100	0 (0)	100	3 (0.2)	100
> 300,000	0 (0)	100	0 (0)	100	0 (0)	100	1 (0.1)	100
Overall frequency	587	100	559	100	561	100	1402	100

Table 7 gives the corresponding information for TP together with the results for cytology negative, as well as HPV-positive/cytology-negative slides from the ARTISTIC study.[17] Of the inadequate slides, 43.3% had a cell count below 5000, compared with 1.7% (42 out of 2552) of adequate slides reported as negative or low-/high-grade dyskaryosis. Almost one-third of inadequate slides had a cell count above 10,000, whereas only 4% (113 out of 2856) of adequate slides reported as negative or low-/high-grade dyskaryosis were below 10,000.

The mean cell count for the ARTISTIC study[17] (known cytological-negative and HPV-positive) slides was 52,166 (34,453), significantly higher than high-grade dyskaryosis (p-value < 0.001) and negative (p-value $= 0.017$) slides, and non-significantly higher than low-grade dyskaryosis (p-value $= 0.507$). The proportion of ARTISTIC study slides that had a cellularity below 5000 (1.3%) was similar to the proportion of dyskaryotic or negative slides. There was, therefore, no indication from these data that potentially false-negative cytology was related to low cellularity. The data also suggest that any threshold of adequate cellularity would apply to HPV positive/cytology negative as it applies to cytology negative/HPV unknown.

Comparison of cellularity of inadequate slides by laboratory

It was relevant to investigate whether or not laboratories used different thresholds for adequate cellularity; therefore, the distribution of cellularity of inadequate and negative slides was considered.

The use of a different threshold could be indicated by the difference between laboratories in the mean cellularity of inadequate slides. This could also be indicated by the differences in the proportion of inadequate slides above the 15,000 cells for SP or 5000 cells for TP MACCs. *Table 8* gives the mean cellularity of inadequate slides and the proportion of inadequate slides with cellularity above these limits for both LBC systems. For SP laboratories, the mean cellularity of inadequate slides ranged from 8266 for laboratory WW up to 24,386 for laboratory YE. The proportion of slides with cellularity above 15,000 ranged from 5% for laboratory WW up to 55% for laboratory YE. For TP laboratories, the mean cellularity of inadequate slides ranged from 4346 for laboratory FF up to 20,780 for laboratory AM. The proportion of slides with cellularity above 5000 ranged from 35% for laboratory HN up to 85% for laboratories IO and KK. These differences are large, the reason for which is unclear.

A one-way ANOVA comparing mean cellularity between laboratories for inadequate slides indicated a significant difference between SP laboratories (F-test p-value $= 0.034$) but not between TP laboratories (p-value $= 0.143$). When the proportion of slides with cellularity above the cut-off points 15,000 and 5000 was compared between laboratories, this revealed a statistically significant difference between laboratories for both SP (chi-squared p-value $= 0.007$) and TP (p-value $= 0.013$). There was, therefore, some evidence of differences between laboratories in the cellularity and the proportion of counts above the current threshold. Formal statistical testing is difficult to interpret when comparing large numbers of units because of the large number of post-hoc pairwise comparisons that can be made. An alternative measure of heterogeneity is to calculate the ICC in order to assess the magnitude of variation between laboratories in either the mean cellularity of inadequate slides or the proportion of inadequate slides above the MACC threshold when compared with the total variation. The ICC estimates the proportion of the total variation that occurs between laboratories. In this context the ICC of 0 indicates that there is no variation between laboratories above that resulting from sampling variation. For SP laboratories, the ICC was 0.031 for the total cell counts and 0.040 for the proportion above the threshold. For TP laboratories, the corresponding figures were 0.027 and 0.032. *Table 8* gives, for each LBC method, the cellularity distribution for slides classified as inadequate by the reading laboratory. For both LBC methods the low ICC values (≈ 0) indicate that any variation within the cell counts is a result of differences between the individual slides and not between laboratories; that is, laboratories are not considered to be producing systematically different cell counts.

TABLE 7 Distribution of cellularity by original slide diagnosis (TP)

	Inadequate		High grade		Mild dyskaryosis		Negative		Cytological negative/HPV positive[a]	
Cellularity	Frequency (%)	Cumulative %	Frequency (%)	Cumulative %	Frequency (%)	Cumulative %	Frequency (%)	Cumulative %	Frequency (%)	Cumulative %
0–2499	139 (22.3)	22.3	3 (0.5)	0.5	3 (0.5)	0.5	5 (0.4)	0.4	2 (0.2)	0.2
2500–4999	131 (21)	43.3	4 (0.7)	1.2	6 (1.1)	1.6	21 (1.5)	1.8	13 (1.1)	1.3
5000–7499	80 (12.8)	56.2	9 (1.6)	2.8	9 (1.6)	3.2	22 (1.6)	3.4	20 (1.7)	3.0
7500–9999	67 (10.8)	66.9	21 (3.7)	6.5	2 (0.4)	3.6	34 (2.4)	5.8	24 (2)	5.0
10,000–14,999	78 (12.5)	79.5	40 (7)	13.5	24 (4.3)	7.8	72 (5.1)	10.9	65 (5.4)	10.4
15,000–19,999	41 (6.6)	86	42 (7.3)	20.8	33 (5.9)	13.7	105 (7.4)	18.3	83 (6.9)	17.3
20,000–24,999	24 (3.9)	89.9	40 (7)	27.8	36 (6.4)	20.1	104 (7.3)	25.6	83 (6.9)	24.2
25,000–29,999	15 (2.4)	92.3	44 (7.7)	35.5	36 (6.4)	26.5	112 (7.9)	33.5	73 (6.1)	30.3
30,000–39,999	18 (2.9)	95.2	88 (15.4)	50.9	89 (15.8)	42.4	178 (12.6)	46	163 (13.6)	43.9
40,000–49,999	18 (2.9)	98.1	93 (16.3)	67.1	83 (14.8)	57.1	171 (12.1)	58.1	137 (11.4)	55.3
50,000–74,999	7 (1.1)	99.2	124 (21.7)	88.8	134 (23.8)	81.0	327 (23.1)	81.2	277 (23.1)	78.4
75,000–99,999	1 (0.2)	99.4	40 (7)	95.8	67 (11.9)	92.9	168 (11.8)	93	146 (12.2)	90.6
100,000–149,999	4 (0.6)	100	23 (4)	100	35 (6.2)	99.1	79 (5.6)	98.6	97 (8.1)	98.7
150,000–199,999	0 (0)	100	1 (0.2)	100	4 (0.7)	100	13 (0.9)	100	16 (1.3)	100
200,000–249,999	0 (0)	100	0 (0)	100	1 (0.2)	100	5 (0.4)	100	0	100
250,000–299,999	0 (0)	100	0 (0)	100	0 (0)	100	1 (0.1)	100	1 (0.1)	100
> 300,000	0 (0)	100	0 (0)	100	0 (0)	100	1 (0.1)	100	0	100
n	623	100	572	100	562	100	1418	100	1200	100

a From the ARTISTIC study.

TABLE 8 Cellularity of inadequate slides for each LBC system and proportion above a MACC

SP					TP				
Laboratory ID	Number of inadequate slides	Mean (SD)	Median	> 15,000, *n* (%)	Laboratory ID	Number of inadequate slides	Mean (SD)	Median	> 5000, *n* (%)
MM	20	14,670 (12,022)	10,916	8 (40)	AA	20	14,109 (23,688)	9081	15 (75)
MS	20	18,725 (14,605)	15,229	10 (50)	AG	39	12,005 (23,898)	4905	19 (48.7)
NN	20	13,176 (22,114)	7447	4 (20)	AM	26	20,780 (25,892)	8418	17 (65.4)
NT	20	10,539 (5240)	10,803	3 (15)	BB	32	9519 (11,757)	4843	16 (50)
OO	20	22,901 (34,454)	10,946	8 (40)	BH	20	7942 (7550)	6231	11 (55)
OU	20	15,053 (23,026)	9910	3 (15)	BN	23	9479 (7891)	5966	16 (69.6)
PP	20	21,523 (33,371)	11,982	6 (30)	CC	22	7878 (12,017)	3646	9 (40.9)
PV	20	9979 (8566)	8650	2 (10)	CI	22	8940 (8174)	6960	13 (59.1)
QQ	20	12,088 (6358)	11,030	5 (25)	CO	20	7247 (5822)	5435	10 (50)
QW	20	17,286 (16,917)	11,982	7 (35)	DD	20	13,327 (17,199)	10,142	12 (60)
RR	22	12,137 (7627)	10,302	8 (36.4)	DJ	25	8843 (7674)	5568	15 (60)
RX	20	13,632 (10,239)	12,122	5 (25)	DP	20	8311 (9379)	5150	11 (55)
SS	20	10,821 (9889)	8371	4 (20)	EE	20	15,657 (17,255)	9214	11 (55)
SY	20	11,724 (8048)	10,533	4 (20)	EK	20	14,782 (13,472)	9810	16 (80)
TT	20	9084 (5257)	10,022	3 (15)	FF	21	4346 (4549)	3182	8 (38.1)
TZ	20	12,595 (7504)	10,806	5 (25)	FL	31	13,118 (22,777)	7556	16 (51.6)
UA	20	11,026 (5126)	9882	3 (15)	GG	20	8630 (8976)	5767	14 (70)
UU	20	13,279 (7277)	12,682	8 (40)	GM	20	13,742 (19,427)	5435	11 (55)
VB	20	19,283 (21,930)	11,118	5 (25)	HH	21	10,208 (10,916)	4110	10 (47.6)
VV	33	13,981 (14,596)	9462	10 (30.3)	HN	20	6521 (11,684)	3115	7 (35)

SP					TP				
Laboratory ID	Number of inadequate slides	Mean (SD)	Median	> 15,000, *n* (%)	Laboratory ID	Number of inadequate slides	Mean (SD)	Median	> 5000, *n* (%)
WC	20	12,722 (16,669)	7559	3 (15)	II	20	11,692 (12,724)	9214	14 (70)
WW	20	8266 (5459)	7335	1 (5)	IO	20	13,746 (8793)	12,196	17 (85)
XD	20	19,639 (14,684)	14,193	9 (45)	JJ	20	12,894 (10,819)	10,804	13 (65)
XX	20	9921 (5231)	9147	2 (10)	JP	20	6040 (5513)	4242	8 (40)
YE	20	24,386 (23,382)	17,156	11 (55)	KK	20	13,993 (12,353)	10,804	17 (85)
YY	32	11,550 (7563)	9910	6 (18.8)	KQ	20	9563 (11,203)	5435	11 (55)
ZF	20	13,008 (7626)	10,974	5 (25)	LL	20	10,584 (13,330)	4574	8 (40)
ZZ	20	10,080 (16,581)	5319	2 (10)	LR	21	7712 (15,271)	4110	8 (38.1)
Total	587	13,980 (15,586)	10,302	150 (25.6)	Total	623	10,870 (14,825)	5833	353 (56.7)
ICC[a]		0.031		0.040	ICC		0.027		0.032

a Intraclass correlation coefficient here based on square root transformation of cellularity.

Comparison of the cellularity of negative slides by laboratory

The use of a different threshold for adequate cellularity by different laboratories could result in varying proportions of negative slides having a cell count below 15,000 cells for SP or 5000 cells for TP. Expected cell counts were small; therefore, a Fisher's exact test indicated that there was no significant relationship (*p*-values 0.335 and 1.000) between the laboratory and the reporting of a negative slide with cellularity below 15,000 and 5000 for SP and TP, respectively. As shown in *Table 9* for SP, the proportion of negative slides with cellularity below 15,000 was 0% in 13 laboratories, 2% in another 13, 4% in one and 10% in the remaining one. A similar profile was seen for TP below a cellularity of 5000. The ICC for TP laboratories and SP laboratories, respectively, was 0.01 and less than 0.001. As with the corresponding ICC for cell counting, these ICCs for the two LBC methods are considered low, indicating that a low proportion of the total variation is a result of between-laboratory variations. The slightly elevated ICC for SP is likely to be explained by laboratory PV which submitted five slides with a cellularity of less than 15,000. The total cell count for these slides revealed that all five slides had a cell count above 10,000; therefore, this could be explained by measurement error in the total cell count.

Key findings

1. The mean counts for inadequate samples at around 14,000 and 11,000 for SP and TP, respectively, are lower than those for negative, mild dyskaryosis and high-grade dyskaryosis samples, which are around 50,000 for both SP and TP.
2. A MACC cut-off point at 15,000 for SP would include only around 25% of the slides reported as inadequate, as well as 97.5%, 97.7% and 98.6% of the slides reported as high-grade dyskaryosis, low-grade dyskaryosis and negative, respectively. This suggests that a SP MACC cut-off of 15,000 could be confirmed as a sensible count reflecting current reporting practices. With regard to TP, however, the MACC recommendation of 5000 would include 56.7% of inadequate samples, and 98.2%, 98.4% and 98.7% of slides reported as high-grade dyskaryosis, low-grade dyskaryosis and negative, respectively. Even above a MACC cut-off of 10,000, which is used in Scotland, 32% of slides reported as inadequate would be included along with 93.5%, 96.4% and 94.2% of high-grade dyskaryosis, low-grade dyskaryosis and negative slides.

Discussion

The results of this cell counting exercise reveal a number of useful findings. The first is that, overall, the mean cell counts for slides classified as inadequate are around 14,000 and 11,000 for SP and TP, respectively. These are considerably lower than for slides reported either negative or abnormal, at around 50,000 for both SP and TP. It should be recognised, however, that the SDs around these means are wide. The second finding relates to the relationship between the MACC and the slide results. A MACC cut-off point of 15,000 for SP would include around 25% of the slides reported as inadequate, as well as 97.5%, 97.7% and 98.6% of slides reported as high-grade dyskaryosis, low-grade dyskaryosis and negative, respectively. This suggests that the widely accepted SP MACC of 15,000 is confirmed as a sensible count reflecting current practice.

With regards to TP, recommended count of TBS of 5000 would include more than 50% of the inadequate samples, and 98.8%, 98.4% and 98.7% of high-grade dyskaryosis, low-grade dyskaryosis and negative samples, respectively. These results, which reflect a cross-section of reporting practice in England, also suggest that a MACC for TP of 10,000, as is currently recommended in Scotland, would still exclude a substantial proportion of slides reported as inadequate and it would also exclude slightly more negative slides.
The proportion of TP slides with cellularity below 5000 and 10,000 in this study read as adequate by participating laboratories can be compared with data from the study by McQueen and Duvall.[11] They reported that, among slides read as adequate (normal or dyskaryotic), 2.5% and 6.5% had a cellularity of less than 5000 and 10,000, respectively. Our data for the same categories were somewhat lower, at 0.9% and 4%, respectively. The finding from our data, that 25% of TP slides read as inadequate had a cellularity between 5000 and 10,000, suggests that, by reducing the MACC from 10,000 (as used in Scotland and other laboratories in England), the TP inadequate rate could be lowered from 2.5% in Scotland[18] to below 2%.

TABLE 9 Proportion of negative slides below thresholds of adequate cellularity

SP			TP		
Laboratory	≤ 15,000, *n* (%)	Number of slides	Laboratory	≤ 5000, *n* (%)	Number of slides
MM	1 (2)	50	AA	0 (0)	50
MS	1 (2)	50	AG	2 (3.8)	53
NN	1 (2)	50	AM	2 (3.9)	51
NT	0 (0)	50	BB	2 (3.8)	53
OO	0 (0)	50	BH	2 (4.0)	50
OU	0 (0)	50	BN	1 (2.0)	50
PP	1 (2)	50	CC	1 (2.0)	50
PV	5 (10)	50	CI	1 (1.9)	52
QQ	0 (0)	50	CO	0 (0)	50
QW	0 (0)	50	DD	0 (0)	50
RR	1 (2)	50	DJ	2 (3.9)	51
RX	2 (4)	50	DP	1 (2.0)	50
SS	0 (0)	50	EE	1 (2.0)	50
SY	1 (2)	50	EK	0 (0)	50
TT	0 (0)	50	FF	2 (4.0)	50
TZ	1 (2)	50	FL	3 (5.3)	57
UA	1 (2)	51	GG	0 (0)	50
UU	0 (0)	50	GM	0 (0)	50
VB	1 (2)	50	HH	1 (2.0)	50
VV	0 (0)	50	HN	1 (2.0)	50
WC	1 (2)	50	II	0 (0)	50
WW	1 (2)	50	IO	0 (0)	50
XD	1 (2)	50	JJ	0 (0)	51
XX	0 (0)	50	JP	1 (2.0)	50
YE	0 (0)	50	KK	0 (0)	50
YY	1 (2)	51	KQ	0 (0)	50
ZF	0 (0)	50	LL	2 (4.0)	50
ZZ	0 (0)	50	LR	1 (2.0)	50
Total	20 (1.4)	1,402	Total	26 (1.8)	1418
Rho	0.01		Rho	< 0.001	

The TP slides included HPV-positive cytologically negative samples from the ARTISTIC study.[17] Looking to the future, this category will constitute the large majority of cytology which will continue to be used to triage women who screen HPV positive. It is clear from these results that the cellularity of slides reported as negative in women whose HPV status is unknown is almost identical to that found in the ARTISTIC slides. This suggests that the results of this study will be applicable to a possible future role of cytology based on HPV status. Consideration is also given to an alternative MACC of 10,000 for TP in the dilution study that follows, when detection rates are analysed.

Objective 4: to assess the impact of varying the cellularity on the likelihood of detection of cytological abnormalities

Introduction

The total number of squamous cells in a cervical sample may influence the probability of detection of abnormality by screeners. To investigate this, serial dilution was used to produce slides from established cases of high- or low-grade abnormality. These slides are referred to as unmixed dilutions. In addition, the relative proportion of dyskaryotic cells compared with normal squamous cells may also influence the probability of detection of abnormality by screeners. This was investigated by producing slides with varying ratios of dyskaryotic to total squamous cells but with a similar background cellularity, which are referred to as mixed dilutions. The unmixed and mixed dilutions can both be thought of as true-positive cases that should be detected by screening as either low- or high-grade dyskaryosis.

Methods

A total of 176 SP and 176 TP cases were selected from material routinely accessioned at the Royal Liverpool University Hospital and the Manchester Cytology Centre that displayed a range of histologically confirmed low- and high-grade cytological abnormalities.

Figure 1 outlines the slide preparation structure for both LBC methods. Diluted preparations were made from each source sample. Seven serial dilutions were made from half of the cases and referred to as 'unmixed dilutions'. The range of dilutions from a cellularity of 5000–10,000 to over 55,000 was skewed towards preparations of lower cellularity, as these were expected to have higher false-negative rates. The remaining half of the cases were serially mixed with known negative cases in varying proportions to establish sets of slides containing different numbers of abnormal cells ranging from < 25 to over 1600. These slides are mixed with normal samples in order to dilute the abnormal cells with normal cells and were, therefore, referred to as 'mixed dilutions' (see *Appendix 6*). In total, 2400 new slides were prepared for each LBC system.

Morphological assessment

To reduce screener outcome bias, the prepared slides were then combined with 1000 negative and 1000 inadequate cases of similar cellularity for each LBC system. Batches of 100 slides were prepared such that each contained similar proportions of unmixed dilutions, mixed dilutions, negative and inadequate slides. Batches were then randomly ordered by source slide type (unmixed dilutions, mixed dilutions, negative and inadequate slides) before being divided into batches that were then assigned to participating laboratories using a bespoke data preparation routine written in the statistical computer package Stata 13 (StataCorp LP, College Station; TX, USA; 2013). Laboratories were asked to screen each slide once under routine primary screening conditions and to assign each slide as inadequate, negative, low-grade dyskaryosis or high-grade dyskaryosis. Each batch of slides was subjected to three independent reviews from three different laboratories resulting in approximately 15,000 slide assessments performed for each LBC system.

Cell counting

Each of the prepared slides (unmixed dilution and mixed dilution) was assessed by one member of a four-panel expert cytology group, resulting in a total cell count for each slide. The total cell count was carried out according to the SOP previously described (see *Appendix 3*).

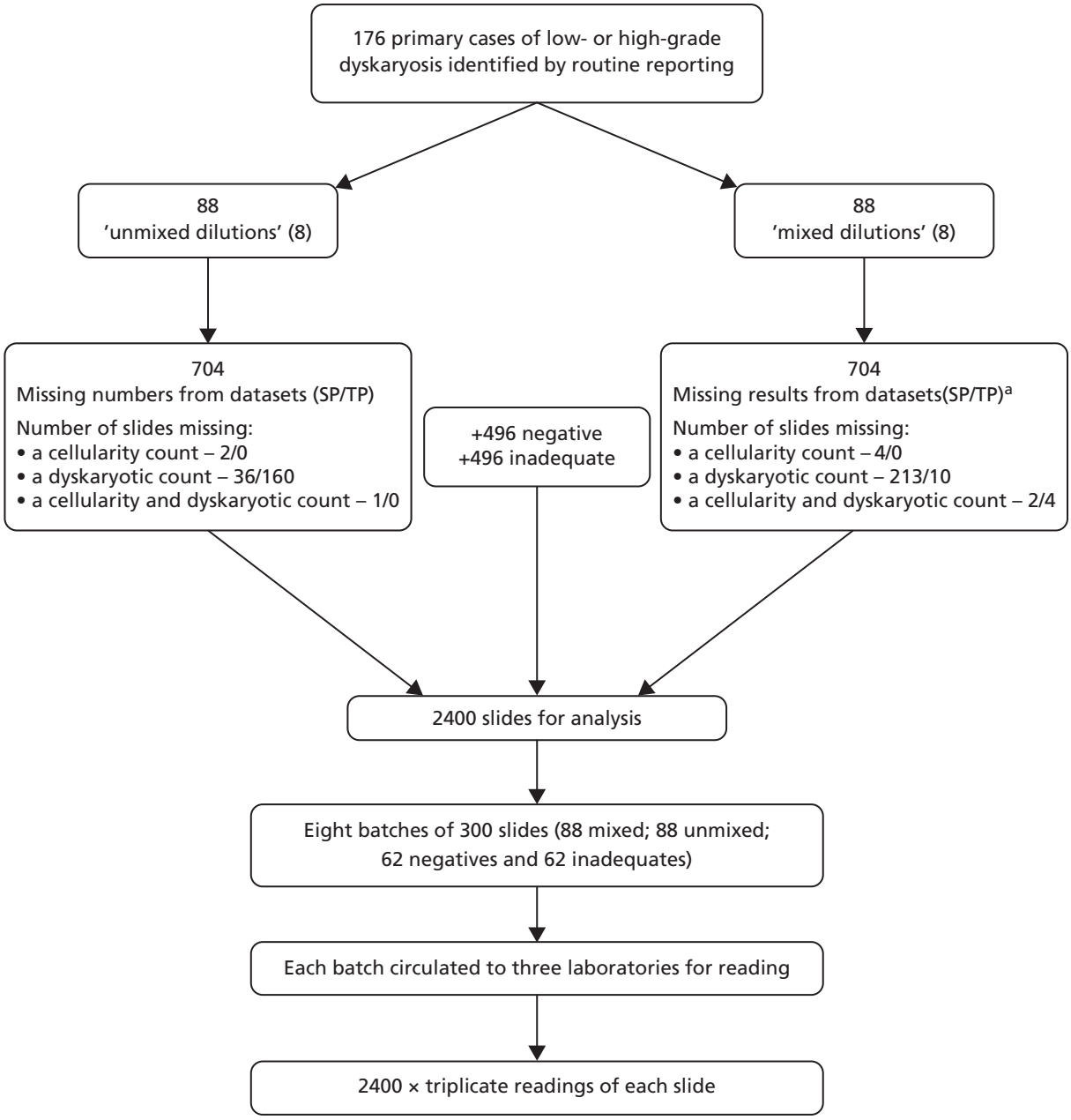


FIGURE 1 Flow diagram outlining slide preparation structure for each LBC method SP/TP. a, Slides with no dyskaryotic cells were removed and replaced. These replaced slides were removed from the main analysis because of a coding error.

A separate count of dyskaryotic cells was also carried out on the unmixed- and mixed-dilution slide preparations. Each slide was initially examined using the same 10-count technique as described for the total specimen cellularity, with the addition that each slide was examined in all four starting positions, that is, starting at 12, 3, 6 and 9 o'clock and heading to the centre. If this 10-count from all four starting positions resulted in no dyskaryotic cells observed, then the whole overall slide was viewed and a count was taken of dyskaryotic cells where present (see *Appendix 5*).

Statistical methods

The reliability of the morphological assessment between the three independent slide reviews was evaluated using a multiassessor kappa coefficient.[19] The reliability of the assessments (inadequate, negative, low or high grade, and high grade) was assessed using the binary scale kappa coefficient.

The overall agreement between the three assessments for the four-category scale of inadequate, negative, low grade and high grade was assessed using the nominal scale kappa coefficients.

To investigate the effect of cellularity on the detection of abnormality, rates for 'low or high' grade and for high grade were calculated for bands of cellularity. A logistic regression model was used to compare the proportion of assessments detected as 'low or high' grade for different ranges of cellularity. Each slide had three morphological assessments, which cannot be considered to be statistically independent. If statistical analysis does not account for the lack of independence, statistical inference will be biased. A modified form of logistic regression called a logistic generalised estimating equation (GEE) regression[20] was, therefore, used. For each band of cellularity, CIs for the detection rate were determined using the robust standard error estimates. It was hypothesised that the detection rate would be lower in slides that were less cellular and also in slides with a large number of cells (high cellularity), making dyskaryotic cells more difficult to identify. To test the hypothesis that the detection rate would decrease in slides thought to be hypocellular or hypercellular, the squamous cell count was fitted as a categorical covariate with three categories representing slides containing the lowest 10%, the highest 10% and the middle 80% of cellular material.

As it is also important to determine the likelihood of a true non-negative slide being defined as non-negative given a change in cellularity, a multilevel logistic regression was repeated for slides assessed to be 'negative' versus 'non-negative'. In this case, the dependent variable is defined as 'non-negative' = 1 and 'negative' = 0. This would mean that an odds ratio (OR) greater than 1 would indicate an increase in the likelihood of a 'non-negative' assessment given an increase in the predictor of interest (cellularity).

To determine the laboratories' ability to detect an abnormal slide (low or high grade) given varying proportions of dyskaryotic cell counts to total cells count, the detection rate within mixed-dilution slides was reported in a cross-tabulation for each combination of the categorised total and categorised dyskaryotic cell counts. To evaluate how change in total cell count relative to dyskaryotic cell count can affect the detection of an abnormal slide, a further multilevel logistic regression was fitted. Here, total cell count and dyskaryotic cell count were both fitted as predictors in the model in the form of an ordinal categorical variable. To determine if the influence of total cell count and dyskaryotic cell count were independent, an interaction variable was also included. If the interaction variable was non-significant, the total cell count and the dyskaryotic cell count would affect the rate of detection independently of each other.

Results

A total of 2400 samples from one of four original sources (inadequate, negative, unmixed dilutions and mixed dilutions) were sent to each of 3 out of 24 laboratories for a morphological assessment – high-grade dyskaryosis, low-grade dyskaryosis, negative and inadequate – resulting in 7200 results.

Agreement of morphological assessment

Table 10 summarises the reliability of the morphological classification by the laboratories giving the multiassessor kappa coefficients across all four categories of high-grade dyskaryosis, low- or high-grade dyskaryosis, negative and inadequate. This is based on the complete data set, that is, the prepared unmixed and mixed dilutions as well as the inadequate and negative slides. The CIs are narrow for all values, as estimates are based on a large sample (2400 slides with 7200 assessments). The overall kappa coefficient was 0.593 (0.571 to 0.610) for SP and 0.609 (95% CI 0.589 to 0.633) for TP. For both LBC systems, classification of 'inadequate' or 'low- or high-grade dyskaryosis' gave higher values of kappa than 'negative' or 'high-grade dyskaryosis'. These differences in kappa coefficients are statistically significant, as there is no overlap of the CIs (*Table 10*). An overall kappa value does not measure which categories were being confused. For example, confusion between negative and high-grade dyskaryosis might be considered more serious than confusion between low- and high-grade dyskaryosis.

Cell counting

Table 11 gives the cellularity for the four types of samples (unmixed dilutions, mixed dilutions, negative and inadequate) that made up the data set. Missing data were the result of a problem with data linkage between morphological cell count data and the data for squamous and dyskaryotic cell counts. The following analysis assesses the agreement between the three observations. Of note, the SP and TP final data sets were missing nine and six total cell counts, respectively. For some unmixed- and mixed-dilutions slides, it was not possible to complete dyskaryotic cell counts. *Table 11* also gives the number of slides in which no dyskaryotic cell could be detected in a detailed count of each slide. In the unmixed dilutions, 64 out of 665 (9.6%) and 47 out of 544 (8.6%) of SP and TP, respectively, were designated hypocellular, that is, below the cellularity of 15,000 and 5000 for SP and TP, respectively. In the inadequate slides, 115 out of 496 (23.2%) and 298 out of 496 (60.1%) were not hypocellular. The results for unmixed dilutions follow in *Objective 4(i): the impact of varying the cellularity on the detection of abnormality using unmixed dilutions* and, for mixed dilutions, in *Objective 4(ii): the impact of varying the relative proportion of abnormal cells on the likelihood of detection of abnormality using mixed dilutions.*

Table 12 shows the comparison between slide classification at the source laboratory and readings from participating laboratories to which slides had been circulated. This shows that almost 25% of inadequate slides were reclassified, the large majority as negative and 5.9% as low or high grade. Among slides originally classified as negative, 2.2% were reclassified as high grade representing a small 'overcall' that is well within acceptable bounds. These observations were similar between SP and TP. The large proportion of mixed dilutions being reported as negative is not unexpected because of the increasing dilution using normal cells.

TABLE 10 Multirater kappa coefficients (95% CI) for morphological assessments

	SP ($n = 2391$ slides)[a]		TP ($n = 2396$ slides)[a]	
Morphological assessment	Kappa	95% CI	Kappa	95% CI
Inadequate	0.732	0.704 to 0.762	0.736	0.707 to 0.762
Negative	0.624	0.600 to 0.649	0.603	0.573 to 0.629
Low or high	0.674	0.649 to 0.694	0.732	0.709 to 0.751
High	0.581	0.547 to 0.613	0.610	0.584 to 0.637
Overall	0.593	0.571 to 0.610	0.609	0.589 to 0.633

a The analysis used morphological assessment of unmixed dilutions, mixed dilutions, negative and inadequate.

TABLE 11 Frequency and distribution of cellularity for all slide types

Type of LBC	Cellularity	Unmixed-dilution frequency (%)	Mixed-dilution frequency (%)	Negative frequency (%)	Inadequate frequency (%)
SP	0–4999	10 (1.4)	–	1 (0.2)	94 (19)
	5000–9999	20 (2.8)	–	–	157 (31.7)
	10,000–14,999	34 (4.8)	2 (0.3)	6 (1.2)	130 (26.2)
	15,000–24,999	119 (16.9)	53 (7.5)	40 (8.1)	66 (13.3)
	25,000–49,999	351 (49.9)	315 (44.7)	236 (47.6)	31 (6.3)
	50,000–74,999	120 (17)	113 (16.1)	144 (29)	13 (2.6)
	75,000+	11 (1.6)	2 (0.3)	69 (13.9)	5 (1)
	Missing	39 (5.5)	219 (31.1)	–	–
	Total	704 (100)	704 (100)	496 (100)	496 (100)
TP	0–2499	11 (1.6)	2 (0.3)	2 (0.4)	100 (20.2)
	2500–4999	36 (5.1)	4 (0.6)	2 (0.4)	98 (19.8)
	5000–7499	35 (5)	11 (1.6)	5 (1)	65 (13.1)
	7500–9999	45 (6.4)	9 (1.3)	16 (3.2)	61 (12.3)
	10,000–14,999	62 (8.8)	33 (4.7)	25 (5)	66 (13.3)
	15,000–24,999	137 (19.5)	86 (12.2)	67 (13.5)	53 (10.7)
	25,000–49,999	146 (20.7)	227 (32.2)	168 (33.9)	44 (8.9)
	50,000–74,999	38 (5.4)	152 (21.6)	114 (23)	6 (1.2)
	75,000+	34 (4.8)	166 (23.6)	97 (19.6)	3 (0.6)
	Missing	160 (22.7)	14 (2.0)	–	–
	Total	704 (100)	704 (100)	496 (100)	496 (100)

TABLE 12 Original slide classification versus the results from three participating laboratories

Result from three participating laboratories	Original slide classification from source laboratory (col%)				
	Inadequate	Negative	Unmixed dilution	Mixed dilution	Total
SP					
Inadequate	1122 (75.4)	36 (2.4)	48 (2.3)	25 (1.2)	1231 (17.1)
Negative	279 (18.8)	1319 (88.6)	196 (9.3)	1143 (54.1)	2937 (40.8)
Low grade	56 (3.8)	100 (6.7)	984 (46.6)	602 (28.5)	1742 (24.2)
High grade	31 (2.1)	33 (2.2)	884 (41.9)	342 (16.2)	1290 (17.9)
Total	1488	1488	2112	2112	7200[a]
TP					
Inadequate	1109 (74.5)	85 (5.7)	51 (2.4)	23 (1.1)	1268 (17.6)
Negative	323 (21.7)	1259 (84.6)	123 (5.8)	398 (18.8)	2103 (29.2)
Low grade	39 (2.6)	111 (7.5)	887 (42)	698 (33)	1735 (24.1)
High grade	17 (1.1)	33 (2.2)	1051 (49.8)	993 (47)	2094 (29.1)
Total	1488	1488	2112	2112	7200[a]

a Three assessments per slide.

Objective 4(i): the impact of varying the cellularity on the detection of abnormality using unmixed dilutions

The analysis of the unmixed-dilution slides investigates the impact of varying the cellularity on the rate of detection of abnormality. *Table 13* compares categorised cellularity and dyskaryotic count for these slides giving frequencies and row percentages. As cellularity reduced with serial dilution, the number of dyskaryotic cells would also be expected to reduce. For both SP and TP the proportion of slides with fewer than 50 dyskaryotic cells did reduce as cellularity increased. A chi-squared test comparing the association of cellularity band with the proportion of slide with less than 50 cells, indicated a possible trend present for both SP (p-value = 0.073) and TP (p-value = 0.002).

The relationship between total cellularity and dyskaryotic count for each of the eight pre-prepared dilutions is given in *Appendix 2*. It can be seen that, with SP, dyskaryotic cells are seen infrequently when the cellularity count is below 10,000, although this degree of cellularity accounted for only 30 out of 665 slides. For TP, 47 out of 544 slides had cell counts below 5000 and 127 out of 544 were below 10,000. Dyskaryotic cells were more frequently seen in TP dilutions, although dyskaryotic cells were infrequent below cell counts of 2500.

TABLE 13 Cellularity and number of dyskaryotic cells for the unmixed dilutions

Type of LBC	Cellularity	Number of dyskaryotic cell (row %)						Total (column %)
		0–	1–	25–	50–	> 100–	Miss	
SP	0–4999	–	1 (10)	–	3 (30)	6 (60)	–	10 (1.4)
	5000–9999	–	4 (19)	3 (14.3)	2 (9.5)	11 (52.4)	1 (4.8)	21 (3)
	10,000–14,999	–	8 (22.2)	9 (25)	5 (13.9)	12 (33.3)	2 (5.6)	36 (5.1)
	15,000–24,999	–	28 (23.3)	17 (14.2)	17 (14.2)	57 (47.5)	1 (0.8)	120 (17)
	25,000–49,999	1 (0.3)	58 (15.4)	60 (15.9)	92 (24.4)	140 (37.1)	26 (6.9)	377 (53.6)
	50,000–74,999	–	19 (15.1)	17 (13.5)	19 (15.1)	65 (51.6)	6 (4.8)	126 (17.9)
	75,000+	–	1 (9.1)	2 (18.2)	5 (45.5)	3 (27.3)	–	11 (1.6)
	Missing	–	–	–	–	2 (66.7)	1 (33.3)	3 (0.4)
	Total	1 (0.1)	119 (16.9)	108 (15.3)	143 (20.3)	296 (42)	37 (5.3)	704 (100)
TP	0–2499	–	9 (69.2)	2 (15.4)	–	–	2 (15.4)	13 (1.8)
	2500–4999	–	20 (40.8)	7 (14.3)	5 (10.2)	4 (8.2)	13 (26.5)	49 (7)
	5000–7499	–	16 (34.8)	5 (10.9)	7 (15.2)	7 (15.2)	11 (23.9)	46 (6.5)
	7500–9999	–	20 (32.3)	9 (14.5)	5 (8.1)	11 (17.7)	17 (27.4)	62 (8.8)
	10,000–14,999	1 (1.2)	25 (30.5)	21 (25.6)	9 (11)	6 (7.3)	20 (24.4)	82 (11.6)
	15,000–24,999	–	57 (32.8)	32 (18.4)	23 (13.2)	25 (14.4)	37 (21.3)	174 (24.7)
	25,000–49,999	–	70 (36.3)	30 (15.5)	24 (12.4)	22 (11.4)	47 (24.4)	193 (27.4)
	50,000–74,999	–	17 (35.4)	12 (25)	6 (12.5)	3 (6.3)	10 (20.8)	48 (6.8)
	75,000+	1 (2.7)	18 (48.6)	6 (16.2)	6 (16.2)	3 (8.1)	3 (8.1)	37 (5.3)
	Missing	–	–	–	–	–	–	–
	Total	2 (0.3)	252 (35.8)	124 (17.6)	85 (12.1)	81 (11.5)	160 (22.7)	704 (100)

Table 13 indicates that SP slides with more than 50,000 cells tend to have more dyskaryotic cells than the corresponding TP slides. Among slides with 50,000 cells, 67% of SP slides had 50 or more dyskaryotic cells, compared with only 19% of TP slides.

Table 14 shows that, among the unmixed dilutions in SP, 28 out of 47 (59.6%) slides classified as inadequate contained fewer than 15,000 cells, compared with 7 out of 185 (3.8%), 66 out of 918 (7.2%) and 91 out of 843 (10.8%) for negative slides, low-grade and high-grade dyskaryosis, respectively. For the TP system, a similar picture emerged, with 20 out of 42 (47.6%) slides classified as inadequate observed in slides with fewer than 5000 cells, compared with 6 out of 108 (5.6%), 50 out of 726 (6.9%) and 65 out of 757 (8.6%) for negative slides and low-grade and high-grade dyskaryosis, respectively. At a threshold of 10,000 for TP, this would have included 33 out of 42 (78.6%) slides classified as inadequate and 17 out of 108 (15.7%), 128 out of 726 (17.6%) and 203 out of 757 (26.8%) for negative slides and low-grade and high-grade dyskaryosis, respectively. These data would suggest that a MACC for TP of 10,000 is excessively high because the range between 5000 and 10,000 included not only 10% of the negatives (11 out of 108) but also 10.7% (78 out of 726) of low grades and 8.2% (138 out of 757) of high grades.

TABLE 14 Morphological assessments by cellularity for the unmixed dilutions

Type of LBC	Specimen cellularity	Morphological assessment frequency (row %)				
		Inadequate	Negative	Low grade	High grade	Total assessments[a]
SP	0–4999	8 (26.7)	–	4 (13.3)	18 (60)	30
	5000–9999	14 (23.3)	1 (1.7)	15 (25)	30 (50)	60
	10,000–14,999	6 (5.9)	6 (5.9)	47 (46.1)	43 (42.2)	102
	15,000–24,999	14 (3.9)	28 (7.8)	149 (41.7)	166 (46.5)	357
	25,000–49,999	5 (0.5)	120 (11.4)	504 (47.9)	424 (40.3)	1053
	50,000–74,999	–	30 (8.3)	183 (50.8)	147 (40.8)	360
	75,000+	–	2 (6.1)	16 (48.5)	15 (45.5)	33
	Missing	1 (0.9)	9 (7.7)	66 (56.4)	41 (35)	117
	Total	48 (2.3)	196 (9.3)	984 (46.6)	884 (41.9)	2112
TP	0–2499	8 (24.2)	–	9 (27.3)	16 (48.5)	33
	2500–4999	12 (11.1)	6 (5.6)	41 (38)	49 (45.4)	108
	5000–7499	5 (4.8)	2 (1.9)	39 (37.1)	59 (56.2)	105
	7500–9999	8 (5.9)	9 (6.7)	39 (28.9)	79 (58.5)	135
	10,000–14,999	3 (1.6)	7 (3.8)	79 (42.5)	97 (52.2)	186
	15,000–24,999	3 (0.7)	25 (6.1)	168 (40.9)	215 (52.3)	411
	25,000–49,999	3 (0.7)	35 (8)	223 (50.9)	177 (40.4)	438
	50,000–74,999	–	2 (1.8)	70 (61.4)	42 (36.8)	114
	75,000+	–	22 (21.6)	58 (56.9)	22 (21.6)	102
	Missing	9 (1.9)	15 (3.1)	161 (33.5)	295 (61.5)	480
	Total	51 (2.4)	123 (5.8)	887 (42)	1051 (49.8)	2112

a Included triplicate readings.

Given that the unmixed-dilution slides contained dyskaryotic cells, *Table 15* indicates a larger number of slides assessed as being 'negative' than we would have expected (SP ≈10% and TP ≈6%). This is a particular concern with regards to SP, as the rate of negative slide assessments appears consistent even when dyskaryotic cell counts were greater than 50. In both LBC methods, investigation of those slides containing one or more 'negative' assessments indicated a low level of agreement (43.7% and 44.4%, respectively) between the three laboratories. A comparison of these slides and those assessed as low grade or high grade did not indicate any significantly different defining characteristics. In TP, 30% of the negative slides did appear to originate from the three (out of 24) assessing laboratories; this may be because of the small sample size and the result was not repeated in SP where the negative slides were spread more evenly across the 24 laboratories.

A greater concern may be if the majority (two or three out of three) of assessments classify the slide as being 'negative'. In both LBC methods, of those slides with at least one 'negative' assessment, approximately 25% included two or more 'negative' assessments, which equates to 4.8% and 3.1% of all unmixed-dilution slides, for each LBC method, respectively. For TP, of those slides with two or more 'negative' assessments, 14% had more than 25 dyskaryotic cells and yet for SP the equivalent was 73%. As seen in TP, we would expect the number of slides identified as negative to drop as the dyskaryotic cell count increased; SP, however, appears to hold constant. The greater tendency of SP to be classified as negative when dyskaryotic cells are present, and the constant rate of negative assessments even when dyskaryotic cells increase, almost certainly reflects the fact that dyskaryotic cells in SP preparations are often in crowded groups, whereas those in TP are more typically dispersed singly. The SP dilutions often contained just a single group of abnormal cells, but this could be made up of tens/hundreds of individual cells. These groups are often more difficult to interpret, as the cells may be incompletely or poorly visualised and are typically of slightly smaller size than cells displayed singly. Thus, a whole group can be missed during screening (identification failure) or can be misinterpreted as benign (interpretation failure).

TABLE 15 Morphological assessments by dyskaryotic cell count for the unmixed dilutions

		Morphological assessment frequency (row %)				
Type of LBC	Dyskaryotic count	Inadequate	Negative	Low grade	High grade	Total assessments
SP	0–24	10 (2.8)	51 (14.2)	216 (60.0)	83 (23.1)	360
	25–49	4 (1.2)	46 (14.2)	178 (54.9)	96 (29.6)	324
	50–99	10 (2.3)	40 (9.3)	235 (54.8)	144 (33.6)	429
	100	23 (2.6)	50 (5.9)	289 (32.5)	520 (59.0)	888
	Missing	1 (0.9)	9 (7.7)	66 (56.4)	41 (35)	117
	Total	48 (2.3)	196 (9.3)	984 (46.6)	884 (41.9)	2112
TP	0–24	26 (3.4)	77 (10.1)	392 (51.4)	267 (35)	762
	25–49	9 (2.4)	20 (5.4)	174 (46.8)	169 (45.4)	372
	50–99	4 (1.6)	6 (2.4)	86 (33.7)	159 (62.4)	255
	100	3 (1.2)	5 (2.1)	74 (30.5)	161 (66.3)	243
	Missing	9 (1.9)	15 (3.1)	161 (33.5)	295 (61.5)	480
	Total	51 (2.4)	123 (5.8)	887 (42)	1051 (49.8)	2112

Detection of abnormal cytology

Table 16 gives the distribution of 'low- or high-grade' assessments and the detection rate stratified by specimen cellularity for the unmixed dilutions. With each slide assessed by three laboratories, each could be 0, 1, 2 or 3 assessments as 'low or high grade'. Suppose the frequency of 1, 2 or 3 positive assessments for a particular stratum of cellularity are $f1$, $f2$, and $f3$. *Table 16* gives these frequencies for each strata. The detection rate is, therefore,

$$\frac{f1 + 2 \times f2 + 3 \times f3}{3 \times n} \tag{2}$$

where n is the number of slides in the strata. For example, if we consider the band of cellularity 0–4999 for SP in *Table 16*, there are 10 slides, of which one slide had no assessments as 'low or high grade', one slide had one 'low- or high-grade' assessment, three slides had two assessments, and five slides were assessed as 'low or high grade' by all three laboratories. These frequencies combine to give a detection rate for 'low or high grade' for this stratum equal to

$$\frac{1 + 2 \times 3 + 3 \times 5}{3 \times 10} = 0.733. \tag{3}$$

TABLE 16 Distribution and rate of abnormal by cellularity for unmixed dilutions

Type of LBC	Cellularity	Low- or high-grade abnormality						
		Number of positive assessments (row %)				Overall detection rate	95% CI	Total no. slides
		*f*0	*f*1	*f*2	*f*3			
SP	0–4999	1 (10)	1 (10)	3 (30)	5 (50)	0.733	0.502 to 0.883	10
	5000–9999	2 (10)	2 (10)	5 (25)	11 (55)	0.750	0.592 to 0.861	20
	10,000–14,999	1 (2.9)	–	9 (26.5)	24 (70.6)	0.882	0.780 to 0.941	34
	15,000–24,999	3 (2.5)	10 (8.4)	13 (10.9)	93 (78.2)	0.882	0.834 to 0.918	119
	25,000–49,999	10 (2.8)	15 (4.3)	65 (18.5)	261 (74.4)	0.881	0.855 to 0.903	351
	50,000–74,999	2 (1.7)	2 (1.7)	20 (16.7)	96 (80)	0.917	0.874 to 0.946	120
	75,000+	–	–	2 (18.2)	9 (81.8)	0.939	0.724 to 0.989	11
	Missing	–	2 (5.1)	6 (15.4)	31 (79.5)	–	–	39
	Total	19 (2.7)	32 (4.5)	123 (17.5)	530 (75.3)	0.844	0.866 to 0.900	704
TP	0–2499	–	3 (27.3)	2 (18.2)	6 (54.5)	0.758	0.551 to 0.888	11
	2500–4999	1 (2.8)	4 (11.1)	7 (19.4)	24 (66.7)	0.833	0.734 to 0.901	36
	5000–7499	–	1 (2.9)	5 (14.3)	29 (82.9)	0.933	0.850 to 0.972	35
	7500–9999	2 (4.4)	3 (6.7)	5 (11.1)	35 (77.8)	0.874	0.792 to 0.927	45
	10,000–14,999	0 (0)	1 (1.6)	8 (12.9)	53 (85.5)	0.946	0.893 to 0.974	62
	15,000–24,999	1 (0.7)	5 (3.6)	15 (10.9)	116 (84.7)	0.932	0.897 to 0.956	137
	25,000–49,999	1 (0.7)	6 (4.1)	23 (15.8)	116 (79.5)	0.913	0.877 to 0.940	146
	50,000–74,999	–	–	2 (5.3)	36 (94.7)	0.982	0.915 to 0.997	38
	75,000+	2 (5.9)	4 (11.8)	8 (23.5)	20 (58.8)	0.784	0.676 to 0.864	34
	Missing	–	5 (3.1)	14 (8.8)	141 (88.1)	–	–	160
	Total	7 (1.0)	32 (4.5)	89 (12.6)	576 (81.8)	0.918	0.902 to 0.931	704

For both SP and TP, the lowest band of cellularity had the lowest rate of detection of 'low- or high-grade' abnormality.

For SP, the detection rate increased from 73% for cellularity between 0 and 4999, to a maximum of 94% for cellularity above 75,000. It is noteworthy that in SP, for slides above the cellularity of 15,000, there was unanimity among the three readers in well over 70% of cases. Similarly, it is of note that, in the case of the 11 slides with cellularity above 75,000, there was unanimity among the three assessors in nine cases. For TP, the detection rate increased from 80% for cellularity between 0 and 2499 to a maximum of 97% at cellularity between 50,000 and 74,999, dropping to 80% for cellularity above 75,000. In TP, unanimity of three readings reached 83% and 85% in slides with cellularity over 5000 and 10,000, respectively. For cellularity between 50,000 and 74,999, there was unanimity of the three assessments for 45 of the 48 slides.

Table 17 presents analyses to investigate the effect of hypo- and hypercellularity on the detection of low-grade or high-grade abnormality. This analysis reports the OR associated with either hypocellular (lowest 10%) or hypercellular (highest 10%) compared with the middle 80% range. The hypocellular threshold at 10% approximated to 5000 and 15,000 for TP and SP, respectively. The hypercellular threshold approximated to 58,000 for both systems. ORs below 1 indicate a reduction in the detection rate compared with the middle 80%. These ORs were estimated using the logistic GEE regression. There was evidence that hypocellularity reduced the detection rate for both LBC systems as the OR of detection was 0.474 (95% CI 0.230 to 0.976; $p = 0.043$) for SP and 0.250 (95% CI 0.108 to 0.577; $p = 0.001$) for TP. The detection rates appear to be increased in the hypercellular band for SP with an OR equal to 3.136 (95% CI 1.199 to 8.199; $p = 0.020$). For TP, the detection rate was reduced (OR 0.366, 95% CI 0.161 to 0.832, $p = 0.016$). The difference between the two systems may be explained by the numbers of dyskaryotic cells in hypercellular slides.

Table 18 summarises the percentage of assessments in each cytological category by cellularity and LBC system for slides prepared as unmixed dilutions. Two ORs were estimated using the logistic GEE regression. The 'non-low or high grade' versus 'low or high grade', and 'non-negative' versus 'negative', were compared with the cellularity cut-offs (15,000 and 5000) for each LBC system. The 'low- or high-grade' percentages for SP are 81.8% for slides below 15,000 and 89.0% for slides above 15,000 (OR 0.56, 95% CI 0.34 to 0.91; $p = 0.020$). Corresponding TP results are in 81.6% of slides assessed as low grade or high grade where cellularity was below 5000, compared with 91.7% for slides above 5000 (OR 0.40, 95% CI 0.23 to 0.71; $p = 0.016$). This indicates that in both systems cellularity greater than 15,000 in SP and 5000 in TP reduces the likelihood of failing to detect 'low- or high-grade' dyskaryosis by approximately 56% and 40% for SP and TP.

The proportion assessed as negative is higher for higher cellularity, with the non-negative compared with negative ORs is 0.410 (95% CI 0.360 to 0.850; $p = 0.021$) and 0.606 (95% 0.217 to 1.695; $p = 0.195$) for SP and TP, respectively. When cellularity is greater than 15,000 in SP and 5000 in TP, the likelihood of a non-negative outcome is reduced by 41% and 61%, respectively.

TABLE 17 Logistic regression estimates of the odds ratio for the effect of hypo- or hypercellularity on the rate of abnormality detection for unmixed dilutions

Type of LBC	Cellularity	Cellularity cut-off (no. slides)[a]	OR	95% CI	*p*-value	Correlation[b]
SP	Hypo (lowest 10%)	< 15,000 ($n = 64$)	0.474	0.230 to 0.976	0.043	0.51
	Hyper (highest 10%)	> 58,000 ($n = 67$)	3.136	1.199 to 8.199	0.020	
TP	Hypo (lowest 10%)	0–5000 ($n = 47$)	0.250	0.108 to 0.577	0.001	0.49
	Hyper (highest 10%)	> 58,000 ($n = 54$)	0.366	0.161 to 0.832	0.016	

a Reference category SP 15,000–58,000 ($n = 534$), TP 5000–58,000 ($n = 443$).
b Correlation is the estimate of the correlation parameter of the exchangeable correlation matrix in the model.

TABLE 18 Morphological assessments by possible cut-off points for unmixed dilution

	< 10,000		10,000–15,000		> 15,000	
SP cellularity	***n* (%)**	**95% CI**	***n* (%)**	**95% CI**	***n* (%)**	**95% CI**
High grade	48 (53.3)	38.6 to 67.5	43 (42.2)	29.2 to 56.3	752 (41.7)	38.5 to 45
Low grade	19 (21.1)	11.6 to 35.3	47 (46.1)	33.1 to 59.7	852 (47.3)	44 to 50.5
Inadequate	22 (24.4)	15.6 to 36.1	6 (5.9)	2.3 to 14.0	19 (1.1)	0.60 to 1.8
Negative	1 (1.1)	0.10 to 10.4	6 (5.9)	2.3 to 14.2	180 (10)	8.5 to 11.7
Number of assessments	90		102		1803	
Number of slides	30		34		601	
	< 5000		**5000–10,000**		**> 10,000**	
TP cellularity	***n* (%)**	**95% CI**	***n* (%)**	**95% CI**	***n* (%)**	**95% CI**
High grade	65 (46.1)	38.0 to 73.8	138 (57.5)	51.2 to 85.4	553 (44.2)	34.7 to 42.8
Low grade	50 (35.46)	28.0 to 55.9	78 (32.5)	26.9 to 46.1	598 (47.8)	37.6 to 46.3
Inadequate	20 (14.18)	9.30 to 24.1	13 (5.42)	3.20 to 9.70	9 (0.72)	0.30 to 1.20
Negative	6 (4.26)	1.90 to 9.90	11 (4.58)	2.60 to 8.60	91 (7.27)	5.00 to 7.60
Number of assessments	141		240		1491	
Number of slides	47		80		497	

Discussion

This prospective study of the effect of serial dilution of samples containing dyskaryotic cells confirmed that there was significant reduction in detection of dyskaryotic cells when squamous epithelial cell counts were below 15,000 and 5000 for SP and TP, respectively. This is the first study to have analysed both LBC systems and utilised a very broad-based evaluation across a large number of cytology laboratories providing services for the cervical screening programmes in the UK. The results support the use of a MACC of 15,000 cells for an adequate SP sample and 5000 cells for an adequate TP sample.

The confidence limits for the detection rates where cellularity is below the MACC are wide because of the small numbers (*Table 16*) and there is some uncertainty regarding the determination of cellularity at the MACC (see *Table 4*).

It has been suggested that SP requires a higher MACC because of the smaller cell deposit area, although this is not a proportionate increase (TP, 15.9 cells/mm^2; SP,112 cells/mm^2), supporting Duval's[21] suggestion that the MACC may be dependent on the preparation method either because TP preferentially enriches and SP depletes the preparation in abnormal cells or because small numbers of abnormal cells are more difficult to detect in SP than in TP preparations. The latter could well be the case because the increased squamous epithelial cell density in SP preparations makes sparse abnormal cells difficult to detect under routine screening conditions.[22]

All of the high-grade cases included in this study were confirmed histologically as CIN 2 or CIN 3; some, but not all, of the low-grade cases had histological confirmation of CIN 1. Participating laboratories were asked to assess slides as inadequate, negative, low- or high-grade dyskaryosis, but no attempt was made to correlate the grade of dyskaryosis with the grade proffered by the submitting laboratory. There are a number of reasons for this decision. Cytological preparations reported as high-grade (severe) squamous dyskaryosis will often contain the complete gamut of changes from low to high grade and the relative proportion of those grades is hugely variable. As a consequence, residual dyskaryosis in highly diluted preparations from high-grade lesions may be of low grade. Hence, when the dilutions were counted,

all dyskaryotic cells were counted, irrespective of their individual grade. This does not detract from the overall purpose of the study, which was to determine what proportion of abnormal cells would be detected, not missed, and either triaged by HPV testing for mild abnormalities or referred directly to colposcopy if high grade.

Each of the outcomes in this study suggests that a MACC of 15,000 and 5000 for SP and TP, respectively, would be associated with greater unanimity of reading and higher detection rates for cytological abnormalities. Raising the MACC for TP to 10,000 would not appear to be relevant according to the results of the unmixed-dilution study.

Objective 4(ii): the impact of varying the relative proportion of abnormal cells on the likelihood of detection of abnormality using mixed dilutions

The analysis of the mixed-dilution slides investigated the impact of the varying proportion of abnormal cells relative to the total cell count on the rate of detection of abnormality by mixing with normal cells and thus diluting the abnormal cells among normal cells. The method for producing the mixed dilutions is reported in *Appendix 6*. *Table 19* compares the categorised cellularity and dyskaryotic count for mixed-dilution slides giving frequencies and row percentages. The mixed dilutions resulted in only 5 out of 704 SP slides

TABLE 19 Cellularity and number of dyskaryotic cells for mixed dilutions

Type of LBC	Cellularity	Number of dyskaryotic cells (row %)						Total (column %)
		0	1	25	50	> 100	Missing	
SP	0–4999	–	–	–	–	–	–	–
	5000–9999	–	–	–	–	–	2 (100)	2 (0.3)
	10,000–14,999	–	–	–	–	2 (66.7)	1 (33.3)	3 (0.4)
	15,000–24,999	7 (11.3)	15 (24.2)	8 (12.9)	11 (17.7)	12 (19.4)	9 (14.5)	62 (8.8)
	25,000–49,999	52 (11.9)	98 (22.5)	45 (10.3)	45 (10.3)	75 (17.2)	121 (27.8)	436 (61.9)
	50,000–74,999	20 (10.8)	28 (15.1)	25 (13.5)	14 (7.6)	26 (14.1)	72 (38.9)	185 (26.3)
	75,000+	–	–	–	1 (10)	1 (10)	8 (80)	10 (1.4)
	Missing	–	1 (16.7)	1 (16.7)	–	2 (33.3)	2 (33.3)	6 (0.9)
	Total	79 (11.2)	142 (20.2)	79 (11.2)	71 (10.1)	118 (16.8)	215 (30.5)	704 (100)
TP	0–2499	1 (50)	1 (50)	–	–	–	–	2 (0.3)
	2500–4999	–	3 (75)	–	–	1 (25)	–	4 (0.6)
	5000–7499	–	6 (54.5)	2 (18.2)	3 (27.3)	–	–	11 (1.6)
	7500–9999	–	6 (66.7)	3 (33.3)	–	–	–	9 (1.3)
	10,000–14,999	1 (3)	12 (36.4)	6 (18.2)	7 (21.2)	7 (21.2)	–	33 (4.7)
	15,000–24,999	2 (2.3)	45 (51.7)	16 (18.4)	5 (5.7)	18 (20.7)	1 (1.1)	87 (12.4)
	25,000–49,999	7 (3)	96 (41.6)	47 (20.3)	32 (13.9)	45 (19.5)	4 (1.7)	231 (32.8)
	50,000–74,999	6 (3.8)	68 (43.6)	29 (18.6)	17 (10.9)	32 (20.5)	4 (2.6)	156 (22.2)
	75,000+	3 (1.8)	88 (52.7)	22 (13.2)	21 (12.6)	32 (19.2)	1 (0.6)	167 (23.7)
	Missing	–	–	–	–	–	4 (100)	4 (0.6)
	Total	20 (2.8)	325 (46.2)	125 (17.8)	85 (12.1)	135 (19.2)	14 (2)	704 (100)

containing cell counts below 15,000 and for TP a similar proportion, 6 out of 704 slides had cell counts below 5000 and 26 out of 704 below 10,000.

Tables 20 and *21* describe the morphological assessments made by each laboratory given the total cellularity count and the dyskaryotic count, respectively. Below SP cell counts of 15,000, there were only 6 out of 1415 assessments, and all were high-grade dyskaryosis. Below TP counts of 5000, there were 18 out of 2070 assessments, including 13 low- or high-grade dyskaryosis. Below TP counts of 10,000, there were 60 out of 2070 assessments, including 43 low- or high-grade dyskaryosis.

Detection of abnormal cytology

Table 22 gives the distribution of 'low- or high-grade' assessments and the detection rate stratified by specimen cellularity for the mixed dilutions. For SP, the detection rate was highest at 58% for cellularity between 15,000 and 24,999, and decreased to 41% for cellularity above 50,000. TP detection rates tended to be higher, at an average of 80%, and increased from 70% for cellularity between 5000 and 7499 to a maximum of 81% for cellularity 50,000–74,999 and 80% for above 75,000. For cellularity below 5000, limited sample sizes of two and four slides meant an inconsistent set of detection rates, hence the large CIs. The same applies to cellularity below 10,000.

TABLE 20 Morphological assessments by cellularity for mixed dilutions

Type of LBC	Specimen cellularity	Morphological assessment frequency (row %)				Total
		Inadequate	Negative	Low grade	High grade	
SP	0–4999	–	–	–	–	–
	5000–9999	–	–	–	–	–
	10,000–14,999	–	–	–	6 (100)	6
	15,000–24,999	7 (4.4)	59 (37.1)	56 (35.2)	37 (23.3)	159
	25,000–49,999	12 (1.3)	459 (48.6)	320 (33.9)	154 (16.3)	945
	50,000–74,999	2 (0.6)	197 (58.1)	97 (28.6)	43 (12.7)	339
	75,000+	0 (0)	5 (83.3)	1 (16.7)	–	6
	Missing	4 (0.6)	423 (64.4)	128 (19.5)	102 (15.5)	657
	Total	25 (1.2)	1143 (54.1)	602 (28.5)	342 (16.2)	2112
TP	0–2499	1 (16.7)	3 (50)	2 (33.3)	–	6
	2500–4499	1 (8.3)	–	3 (25)	8 (66.7)	12
	5000–7499	5 (15.2)	5 (15.2)	7 (21.2)	16 (48.5)	33
	7500–9999	1 (3.7)	6 (22.2)	9 (33.3)	11 (40.7)	27
	10,000–14,999	6 (6.1)	19 (19.2)	23 (23.2)	51 (51.5)	99
	15,000–24,999	2 (0.8)	55 (21.3)	82 (31.8)	119 (46.1)	258
	25,000–49,999	2 (0.3)	122 (17.9)	239 (35.1)	318 (46.7)	681
	50,000–74,999	1 (0.2)	85 (18.6)	146 (32)	224 (49.1)	456
	75,000+	4 (0.8)	95 (19.1)	173 (34.7)	226 (45.4)	498
	Missing	–	8 (19)	14 (33.3)	20 (47.6)	42
	Total	23 (1.1)	398 (18.8)	698 (33)	993 (47)	2112

TABLE 21 Morphological assessments by dyskaryotic cell count for mixed dilutions

		Morphological assessment frequency (row %)				
LBC System	Dyskaryotic count	Inadequate	Negative	Low grade	High grade	Total
SP	0–24	12 (1.8)	466 (70.2)	145 (21.9)	37 (5.6)	663
	25–49	5 (2.1)	96 (40.5)	106 (44.7)	27 (11.4)	237
	50–99	2 (0.9)	71 (33.3)	82 (38.5)	57 (26.8)	213
	100	2 (0.6)	87 (24.6)	141 (39.8)	119 (33.6)	354
	Missing	4 (0.6)	423 (64.4)	128 (19.5)	102 (15.8)	657
	Total	21 (1.0)	728 (34.5)	478 (22.6)	240 (11.4)	2112
TP	0–24	17 (1.6)	304 (29.4)	385 (37.2)	329 (31.8)	1035
	25–49	5 (1.3)	42 (11.2)	137 (36.5)	191 (50.9)	375
	50–99	–	12 (4.7)	75 (29.4)	168 (65.9)	255
	100	1 (0.2)	32 (7.9)	87 (21.5)	285 (70.4)	405
	Missing	–	8 (19.0)	14 (33.3)	20 (47.6)	42
	Total	23 (1.1)	390 (18.5)	684 (32.4)	973 (46.1)	2112

TABLE 22 Distribution and rate of abnormal assessments by cellularity for mixed dilutions

		Low- or high-grade abnormality						
		Number of positive assessments				Overall detection rate	95% CI	Total no. slides
Mixed dilutions only	Cellularity	*f*0	*f*1	*f*2	*f*3			
SP	0–4999	–	–	–	–	–	–	–
	5000–9999	–	–	–	–	–	–	–
	10,000–14,999	–	–	–	2 (100)	1	–	2
	15,000–24,999	13 (24.5)	6 (11.3)	15 (28.3)	19 (35.8)	0.585	0.474 to 0.687	53
	25,000–49,999	99 (31.4)	59 (18.7)	56 (17.8)	101 (32.1)	0.502	0.457 to 0.546	315
	50,000–74,999	44 (38.9)	24 (21.2)	19 (16.8)	26 (23)	0.413	0.341 to 0.488	113
	75,000+	1 (50)	1 (50)	–	–	0.167	0.010 to 0.805	2
	Missing	107 (48.9)	38 (17.4)	30 (13.7)	44 (20.1)	–	–	219
	Total	264 (37.5)	128 (18.2)	120 (17)	192 (27.3)	0.447	0.417 to 0.477	704
TP	0–2499	1 (50)	–	1 (50)	–	0.333	0.048 to 0.833	2
	2500–4999	–	–	1 (25)	3 (75)	0.917	0.407 to 0.994	4
	5000–7499	–	4 (36.4)	2 (18.2)	5 (45.5)	0.697	0.457 to 0.863	11
	7500–9999	1 (11.1)	1 (11.1)	2 (22.2)	5 (55.6)	0.741	0.471 to 0.902	9
	10,000–14,999	3 (9.1)	3 (9.1)	10 (30.3)	17 (51.5)	0.747	0.616 to 0.845	33
	15,000–24,999	8 (9.3)	10 (11.6)	13 (15.1)	55 (64.0)	0.779	0.703 to 0.840	86
	25,000–49,999	15 (6.6)	19 (8.4)	41 (18.1)	152 (67.0)	0.818	0.775 to 0.854	227
	50,000–74,999	11 (7.2)	14 (9.2)	25 (16.4)	102 (67.1)	0.811	0.758 to 0.855	152
	75,000+	12 (7.2)	18 (10.8)	27 (16.3)	109 (65.7)	0.801	0.749 to 0.844	166
	Missing	–	2 (14.3)	4 (28.6)	8 (57.1)	–	–	14
	Total	51 (7.2)	69 (9.8)	126 (17.9)	456 (64.8)	0.800	0.777 to 0.823	704

To investigate how varying the dyskaryotic cell count with total cell count affects the detection of an abnormal slide, *Table 23* gives the detection rates of a 'low- or high'-grade slide given varying total and dyskaryotic cell counts. In addition, rate ratios are given for each combination of total and dyskaryotic cell count when compared with the lowest total cell count group (SP = 15,000/TP = 5000) and the highest dyskaryotic cell count (> 50). A rate ratio greater than 1 indicates an increased rate of detection whereas a rate ratio less than 1 indicates a decrease.

For SP, detection rates appear to decrease as total cell count increases and increase as dyskaryotic cell count increases. This resulted in the largest detection rate (0.768) occurring when total cell count was below 25,000 and dyskaryotic cell count higher than 50, and the smallest detection rate (0.139) when the total cell count was higher than 50,000 and the dyskaryotic cell count lower than 25. TP tended to show an increase in detection as dyskaryotic cell count increased, but stayed constant as total cell count increased.

To investigate if trends were present in the ordinal variables described in *Table 23*, a logistic GEE regression model was fitted to assess the likelihood of detecting a 'low or high grade' versus 'non-low or high grade'. The results in *Table 24* reveal that in both LBC methods the interaction between total and dyskaryotic cell count was fitted and found to be non-significant, indicating that a change in total and dyskaryotic cell counts independently affects the odds of 'low- or high-grade' result, hence the interaction was removed in the subsequent model.

Once the interaction was removed, SP results confirmed a significant decrease in the likelihood of detecting 'low- or high-grade' outcome as cellularity increased (OR 0.512, 95% CI 0.328 to 0.798) and a significant increase as dyskaryotic cell count increased (OR 4.949, 95% CI 3.603 to 6.798). For TP, there was a slight increase in the detection rate as total cellularity increased, although this was not significant (*p*-value 0.214). The detection rate increased with an increase in dyskaryotic cell count (OR 4.315, 95% CI 3.103 to 6.000).

Discussion

Two findings emerge from the mixed-dilution study, neither of which is unexpected. The first is that as cellularity increased, thus diluting the number of dyskaryotic cells, there was a significant decrease in the likelihood of detecting dyskaryotic cells, although in TP this was not statistically significant. The second relates to the number of dyskaryotic cells themselves, and, for both SP and TP, as the dyskaryotic cell count decreased, so did the likelihood of detection. Compared with a reference standard of more than 50 dyskaryotic cells on slides above the MACC, the OR for detection below 25 dyskaryotic cells was 0.49 and 0.74 for SP and TP, respectively. This suggests that a threshold of 25 dyskaryotic cells could be considered a reasonable threshold below which the chance of detecting abnormal cells is significantly reduced. This issue is relevant to medicolegal practice.

TABLE 23 Detection rates of a low- or high-grade morphological result given cellularity and dyskaryotic cell counts (mixed-dilution slides)

		Dyskaryotic cell count grouped						
		0–		25–		50–		Total
LBC system	Cell grouped	Rate (*S*/*n*)	Rate ratio	Rate (*S*/*n*)	Rate ratio	Rate (*S*/*n*)	Rate ratio	Rate (*S*/*n*)
SP	15,000–24,999	0.379 (25/66)	0.49	0.625 (15/24)	0.81	0.768 (53/69)	–	0.585 (93/159)
	25,000–49,999	0.304 (137/450)	0.40	0.585 (79/135)	0.76	0.717 (258/360)	0.93	0.502 (474/945)
	50,000+	0.139 (20/144)	0.18	0.520 (39/75)	0.68	0.651 (82/126)	0.85	0.409 (141/345)
	Total	0.276 (182/660)	–	0.568 (133/234)	–	0.708 (393/555)	–	0.489 (708/1449)
TP	5000–999	0.694 (25/36)	0.69	0.600 (9/15)	0.60	1 (9/9)	–	0.716 (43/60)
	10,000–24,999	0.661 (119/180)	0.66	0.864 (57/66)	0.86	0.892 (99/111)	0.89	0.770 (275/357)
	25,000–49,999	0.696 (215/309)	0.70	0.887 (125/141)	0.89	0.939 (217/231)	0.94	0.818 (557/681)
	50,000–74,999	0.689 (153/222)	0.69	0.908 (79/87)	0.91	0.939 (138/147)	0.94	0.811 (370/456)
	75,000+	0.703 (192/273)	0.70	0.879 (58/66)	0.88	0.937 (149/159)	0.94	0.801 (399/498)
	Total	0.690 (704/1020)	–	0.875 (328/375)	–	0.932 (612/657)	–	0.801 (1644/2052)

n, total number of assessments; *S*, number of successful 'low- or high-grade' assessments.
The rate ratio was compared with baseline rate at cellularity below 15,000/5000 and more than 50 dyskaryotic cells.

TABLE 24 Logistic regression investigation of the detection rates corresponding to the above ordinal cellularity and dyskaryotic variables (mixed dilutions)

LBC method	Low- or high-grade result	OR	*p*-value	95% CI
SP	Ordinal specimen cellularity	0.384	0.005	0.197 to 0.746
	Ordinal dyskaryotic count	3.577	< 0.001	1.919 to 6.667
	Interaction	1.338	0.247	0.817 to 2.192
TP	Ordinal specimen cellularity	1.089	0.506	0.848 to 1.397
	Ordinal dyskaryotic count	3.461	0.001	1.679 to 7.136
	Interaction	1.094	0.509	0.836 to 1.432

The interaction effect is non-significant; therefore, interaction should be ignored. In addition, ordinal variable used to reduce the impact of large outliers in both cellularity and dyskaryotic count.

Chapter 2 General discussion

The findings of this study can be summarised as follows:

(a) As all the SP laboratories surveyed use a MACC of 15,000 cells, TP laboratories vary, with a range of 5000 to 15,000 cells being used.
(b) Cell counting is associated with a moderate degree of interassessor agreement, with only a small proportion of counts showing substantial disagreement between assessors.
(c) When data from a large range of laboratories and assessors are collated, it is clear that a large proportion of slides classified as inadequate are associated with cell counts above the widely used MACC of 15,000 cells for SP and within the range of 5000 to 10,000 cells for TP.
(d) Dilutional studies of samples indicated that:

- Unmixed dilution showed that above MACC thresholds of 15,000 and 5000 cells for SP and TP, respectively, there was a significant increase in unanimity of reporting abnormalities and an increase in the likelihood of detecting dyskaryotic cells.
- Mixed dilutions demonstrated that above the MACC for SP, as cellularity increased, the likelihood of detecting dyskaryotic cells decreased and also that once the dyskaryotic cell count fell below 25 the likelihood of detecting abnormal cells was reduced with both SP and TP.

Criteria for the assessment of adequacy of cervical cytology samples have been widely discussed for many years and remain the subject of debate. In the UK, in common with other countries with established cytology-based cervical screening programmes, it was generally agreed that, for routine cervical screening using conventional Papanicolaou cervical smears, the primary indicator of adequacy was the presence of a sufficient number of squamous epithelial cells, possibly supplemented by morphological indicators of transformation zone sampling, namely mucus, metaplastic squamous epithelial cells and endocervical cells. However, consistent and reliable identification of these criteria has been questioned.[23,24]

Following the introduction of LBC in the USA, TBS required a minimum of 5000 squamous cells on the slide for a LBC preparation to be regarded as adequate and provided comprehensive guidance on how the MACC should be determined.[10] It was recommended that a minimum of 10 microscopic fields, usually at ×40 objective magnification, should be assessed along a diameter that includes the centre of the preparation and an average number of cells per field estimated. It was also recommended that when there are holes or empty areas on the preparation (as is often the case in TP samples), the percentage of the hypocellular areas should be estimated. The fields counted should reflect this proportion, immediately introducing a subjective element into a numerical assessment. A study of the cellularity of liquid-based preparations for normal, abnormal and false-negative cervical cytology cases in which cellular objects were counted using a fully automated microscope. This demonstrated that while the population of abnormal slides tended to have higher cellularity, the population of false-negative slides could not be distinguished by their cellularity. It was concluded that cellularity does not provide assurance of adequacy and recommended that any cellularity criterion should be based on measurement of the prevalence of abnormal cells on abnormal slides.[25] Subsequently, only one study supported a MACC of 5000 cells,[26] and others demonstrated that detection of abnormality increased substantially as cell numbers increased to 10,000[25] or even higher.[12] Unfortunately, the last-mentioned study, which measured the prevalence of abnormal cells on abnormal slides was presented only as a poster at the American Society of Cytopathology and has not been subjected to peer-reviewed publication. Umana *et al.*[27] and others have also recently reported that there was little, if any, difference in the likelihood of abnormal cells being seen in TP slides containing 10–20 cells or > 20 cells per high-power field. They did not find a significant difference in abnormal cells being seen in slides with fewer than 10 cells per high-power field (equivalent to approximately 13,000 cells on a slide). This supported the findings of Bolick[12] and others that abnormalities were less likely to be found in TP slides containing fewer than 20,000 cells.

This current study had a number of strengths and a number of limitations. The strengths were that it was a prospective exercise that involved both SP and TP, that over 50 cervical screening laboratories were involved, and that a standard counting protocol was used. Weaknesses included the small number of hypocellular slides and 'missing' slides. The 'missing' data occurred because a number of slides showing no dyskaryotic cells were removed and replaced. A coding error while labelling the replaced slides meant that we were not confident that we could accurately match the cell counts with the morphological assessments made by the three laboratories. Therefore, we were forced to remove these slides from the analysis. Histology was not used as an end point because the purpose of the study was to address detection of abnormal cells on cytology and the adequacy of slides in terms of cell counts. Despite some shortcomings, this study represents a more rigorous exercise than any other previously undertaken in the UK, and indeed internationally. Since 2008, when this study was initiated, there have not been any key studies which undermine the relevance of our findings; specifically, there has not been another peer-reviewed publication on a dilutional study. Although laboratory practice differs in different countries, the SP and TP LBC systems are now widely employed in the developed world, and the findings, therefore, have international relevance. In particular, the MACC for TP and SP should be considered for adoption into national laboratory practice guidelines. Cell counting is not practical for every slide; however, the standardised counting protocol used in this study for TP and SP is easy to follow and should also be incorporated into national laboratory practice guidelines for perceived low cellularity samples during initial screening.

Acknowledgements

We wish to acknowledge the large number of cytology laboratory staff who co-operated and collaborated with this study, which involved time and effort over and above their routine duties.

Jan Perriton acted as study co-ordinator at Liverpool Women's Hospital.

Jean Mather (Manchester), Chris Evans (Liverpool) and Kay Ellis (Sheffield) provided considerable central laboratory support.

Peter Sasieni served on the Study Management Group and provided critical comment.

We also acknowledge the valuable assistance of Linsey Nelson in preparing this report.

Contribution of authors

Henry C Kitchener (Professor of Gynaecological Oncology) contributed to the study design and drafted the manuscript.

Matthew Gittins (Research Assistant) undertook and co-reported the statistical analysis.

Mina Desai (Consultant Cytopathologist) codesigned the study, contributed to the laboratory project and provided critical comment on the report.

John HF Smith (Consultant Cytopathologist) contributed to the laboratory project, contributed to the drafting of the report and provided critical comment.

Gary Cook (Consultant Epidemiologist) contributed critical comment on the report and served with the other grant holders on the Project Management Team.

Chris Roberts (Professor of Biostatistics) supervised the statistical analysis and co-wrote the statistical report.

Lesley Turnbull (Consultant Cytopathologist) contributed to study design, co-ordinated the laboratory exercises and contributed to the drafting of the report.

References

1. Anderson GH, Boyes DA, Benedet JL, Le Riche JC, Matisic JP, Suen KC, *et al.* Organisation and results of the cervical cytology screening programme in British Columbia, 1955–85. *Br Med J* 1988;**296**:975–8. http://dx.doi.org/10.1136/bmj.296.6627.975
2. Trent Cancer Registry/NHS Cancer Screening Programmes. *Profile of Cervical Cancer in England – Incidence, Mortality and Survival*. Sheffield: Trent Cancer Registry; 2011.
3. National Board of Health, Planning Division. *Cervical Cancer Screening – Guidelines 2007 – Summary*. Copenhagen: National Board of Health, Planning Division; 2007.
4. Dillner J, Rebolj M, Birembaut P, Petry K-U, Szarewski A, Munk C, *et al.* Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: joint European cohort study. *BMJ* 2008;**337**:a1754.
5. Sasieni P, Castanon A, Cuzick J. Effectiveness of cervical screening with age: population based case–control study of prospectively recorded data. *BMJ* 2009;**339**:b2968. http://dx.doi.org/10.1136/bmj.b2968
6. Papanicolaou GN. A new procedure for staining vaginal smears. *Science* 1942;**95**:438–9. http://dx.doi.org/10.1126/science.95.2469.438
7. Moss S, Gray A, Marteau T, Legood R, Henstock E, Maissi E. *Evaluation of HPV/LBC*. London: Cervical Screening Pilot Studies, Department of Health; 2004.
8. Arbyn M, Bergeron C, Klinkhamer P, Martin-Hirsch P, Siebers AG, Bulten J. Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. *Obstet Gynecol* 2008;**111**:167–77. http://dx.doi.org/10.1097/01.AOG.0000296488.85807.b3
9. National Institute for Health and Care Excellence (NICE). *Guidance on the Use of Liquid-Based Cytology for Cervical Screening*. London: NICE; 2003.
10. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, *et al.* The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 2002;**287**:2114–19. http://dx.doi.org/10.1001/jama.287.16.2114
11. McQueen F, Duvall E. Using a quality control approach to define an 'adequately cellular' liquid-based cervical cytology specimen. *Cytopathology* 2006;**17**:168–74. http://dx.doi.org/10.1111/j.1365-2303.2006.00344.x
12. Bolick DR, Staley BE, Ke Lin K. Establishing diagnostic curves to estimate the false negative proportion of paps due to low cellularity. *Acta Cytol* 2002;**46**:965.
13. Siebers AG, van der Laak JA, Huberts-Manders R, Vedder JE, Bulten J. Accurate assessment of cell density in low cellular liquid-based cervical cytology. *Cytopathology* 2013;**24**:216–21. http://dx.doi.org/10.1111/j.1365-2303.2012.00990.x
14. Smith JH. ABC3 Part I: a review of the guidelines for terminology, classification and management of cervical cytology in England. *Cytopathology* 2012;**23**:353–9. http://dx.doi.org/10.1111/cyt.12031
15. Efron B, Tibshirani RJ. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1986;**1**:54–77. http://dx.doi.org/10.1214/ss/1177013815
16. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;**43**:543–9. http://dx.doi.org/10.1016/0895-4356(90)90158-L

17. Kitchener HC, Almonte M, Gilham C, Dowie R, Stoykova B, Sargent A, *et al.* ARTISTIC Trial Study Group. ARTISTIC: a randomised trial of human papillomavirus (HPV) testing in primary cervical screening. *Health Technol Assess* 2009;**13**(51). http://dx.doi.org/10.3310/hta13510

18. Information Services Division. *Scottish Cervical Screening Programme Statistics 2012–13*. Scottish Cervical Screening Programme; 2013.

19. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd edn. New York, NY: Wiley; 2003. http://dx.doi.org/10.1002/0471445428

20. Laing K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;**73**:13–22. http://dx.doi.org/10.1093/biomet/73.1.13

21. Duvall E. ABC3 and LBC – adequate or not? *Cytopathology* 2013;**24**:211–15. http://dx.doi.org/10.1111/cyt.12081

22. Gupta N, John D, Dudding N, Crossley J, Smith JH. Factors contributing to false-negative and potential false-negative cytology reports in SurePath liquid-based cervical cytology. *Cytopathology* 2013;**24**:39–43. http://dx.doi.org/10.1111/j.1365-2303.2012.00992.x

23. Renshaw AA, Friedman MM, Rahemtulla A, Granter SR, Dean BR, Cronin JA, *et al.* Accuracy and reproducibility of estimating the adequacy of the squamous component of cervicovaginal smears. *Am J Clin Pathol* 1999;**111**:38–42.

24. Slater DN, Hewer EM, Melling SE, Rice S. External quality assessment in gynaecological cytology: The Trent Region experience. The Trent Regional Gynaecological Pathology Quality Assurance Group for the National Health Service Cervical Screening Programme. *Cytopathology* 2002;**13**:206–19. http://dx.doi.org/10.1046/j.1365-2303.2002.00391.x

25. Bishop JW. Cellularity of liquid-based, thin-layer cervical cytology slides. *Acta Cytol* 2002;**46**:633–6. http://dx.doi.org/10.1159/000326967

26. Studeman KD, Ioffe OB, Puszkiewicz J, Sauvegeot J, Henry MR. Effect of cellularity on the sensitivity of detecting squamous lesions in liquid-based cervical cytology. *Acta Cytol* 2003;**47**:605–10. http://dx.doi.org/10.1159/000326576

27. Umana A, Dunsmore H, Herbert A, Jokhan A, Kubba A. Are significant numbers of abnormal cells lost on the discarded ThinPrep® broom when used for cervical cytology? *Cytopathology* 2013;**24**:228–34. http://dx.doi.org/10.1111/cyt.12029

Appendix 1 Laboratory questionnaire

Study Site Survey

Q 1. Name of Laboratory:
Hospital:
Trust:

Q 2. Country / Region:

England (please specify region)

- North West
- North East, Yorkshire and The Humber
- South Central
- East of England
- West Midlands
- East Midlands
- South West
- South East
- London

Scotland_____

Wales_____

Q 3. Postal Address of Laboratory:

Road
Town
County
Postal code

Q 4. Postal Address for deliveries if different from above:

Road
Town
County
Postal code

Q 5. Special instructions for deliveries if appropriate:

Q 6. Name and address of Lead Pathologist:

Name
Road
Town
County
Postal code
Email address
Tel no
Fax no

Q 7. Name, title and address of designated laboratory lead for study (Study Liaison Officer):

Name
Road
Town
County
Postal code
Email address
Tel no
Fax no

Q 8. To ensure the smooth running of the study, please provide the name and title of designated deputy for study (Deputy Study Liaison Officer):

Name
Road
Town
County
Postal code
Email address
Tel no
Fax no

Q 9. Type of LBC system used:

Cytyc ThinPrep

SurePath

Both

Q 10. LBC training data

Approximate start date of training

Approximate date of completion of training

Cytology Training Centre responsible for delivery of training

Date of full conversion of laboratory to LBC

Q 11. Provide details of criteria currently used by your laboratory for assessing LBC sample adequacy including determinants of transformation zone sampling (quantative & qualitative):

_____
_____
_____
_____
_____
_____
_____
_____

Q 12. Does the laboratory have SOPs for the assessment? ☐

Q 13. If yes please attach relevant documents including, for example, method for cell counting

Q 14. Please also attach KC61 data for 2006/07

Thank you for taking the time to complete this survey. Pease return the completed form to the North West Quality Assurance Reference Centre at your earliest convenience. See over for full address.
Please note that the trial is not related to QA activity and that individual laboratories will not be identified

Appendix 2 Specimen cellularity compared with the numbers of dyskaryotic cells, split by dilution levels; SurePath (SP) and ThinPrep (TP)

TABLE 25 Specimen cellularity group compared with dyskaryotic group, split by dilution level: SP

SP unmixed dilutions	Specimen cellularity group	Dyskaryotic group					Total
		0	1–	25–	50–	100–	
D1	0–4999	–	–	–	–	1 (100)	1
	5000–9999	–	–	1 (50)	–	1 (50)	2
	10,000–14,499	–	–	–	–	1 (100)	1
	15,000–24,999	–	–	1 (20)	1 (20)	3 (60)	5
	25,000–74,999	1 (1.4)	4 (5.7)	9 (12.9)	16 (22.9)	40 (57.1)	70
	75,000+	–	–	–	–	1 (100)	1
	Total	1 (1.3)	4 (5)	11 (13.8)	17 (21.3)	47 (58.8)	80
D2	0–4999	–	–	–	–	1 (100)	1
	5000–9999	–	–	–	–	–	–
	10,000–14,999	–	–	–	–	1 (100)	1
	15,000–24,999	–	1 (12.5)	3 (37.5)	1 (12.5)	3 (37.5)	8
	25,000–74,999	–	11 (15.9)	10 (14.5)	14 (20.3)	34 (49.3)	69
	75,000+	–	–	–	2 (50)	2 (50)	4
	Total	–	12 (14.5)	13 (15.7)	17 (20.5)	41 (49.4)	83
D3	0–4999	–	–	–	–	–	–
	5000–9999	–	–	–	–	2 (100)	2
	10,000–14,999	–	–	1 (50)	–	1 (50)	2
	15,000–24,999	–	2 (13.3)	1 (6.7)	5 (33.3)	7 (46.7)	15
	25,000–74,999	–	7 (11.7)	11 (18.3)	17 (28.3)	25 (41.7)	60
	75,000+	–	–	–	3 (100)	–	3
	Total	–	9 (11)	13 (15.9)	25 (30.5)	35 (42.7)	82
D4	0–4999	–	–	–	–	–	–
	5000–9999	–	–	–	–	3 (100)	3
	10,000–14,999	–	–	2 (50)	–	2 (50)	4
	15,000–24,999	–	2 (16.7)	4 (33.3)	–	6 (50)	12
	25,000–74,999	–	8 (13.8)	10 (17.2)	14 (24.1)	26 (44.8)	58
	75,000+	–	1 (50)	1 (50)	–	–	2
	Total	–	11 (13.9)	17 (21.5)	14 (17.7)	37 (46.8)	79

continued

TABLE 25 Specimen cellularity group compared with dyskaryotic group, split by dilution level: SP (*continued*)

SP unmixed dilutions	Specimen cellularity group	Dyskaryotic group					Total
		0	1–	25–	50–	100–	
D5	0–4999	–	–	–	–	1 (100)	1
	5000–9999	–	–	–	–	1 (100)	1
	10,000–14,999	–	1 (25)	1 (25)	–	2 (50)	4
	15,000–24,999	–	6 (37.5)	–	3 (18.8)	7 (43.8)	16
	25,000–74,999	–	7 (12.3)	9 (15.8)	13 (22.8)	28 (49.1)	57
	75,000+	–	–	1 (100)	–	–	1
	Total	–	14 (17.5)	11 (13.8)	16 (20)	39 (48.8)	80
D6	0–4999	–	–	–	1 (50)	1 (50)	2
	5000–9999	–	–	–	–	2 (100)	2
	10,000–14,999	–	1 (14.3)	3 (42.9)	1 (14.3)	2 (28.6)	7
	15,000–24,999	–	3 (21.4)	2 (14.3)	1 (7.1)	8 (57.1)	14
	25,000–74,999	–	11 (19.3)	8 (14)	12 (21.1)	26 (45.6)	57
	75,000+	–	–	–	–	–	–
	Total	–	15 (18.3)	13 (15.9)	15 (18.3)	39 (47.6)	82
D7	0–4999	–	–	–	1 (50)	1 (50)	2
	5000–9999	–	–	1 (50)	1 (50)	–	2
	10,000–14,999	–	4 (50)	–	2 (25)	2 (25)	8
	15,000–24,999	–	3 (15)	3 (15)	3 (15)	11 (55)	20
	25,000–74,999	–	12 (23.5)	8 (15.7)	15 (29.4)	16 (31.4)	51
	75,000+	–	–	–	–	–	–
	Total	–	19 (22.9)	12 (14.5)	22 (26.5)	30 (36.1)	83
D8	0–4999	–	1 (33.3)	–	1 (33.3)	1 (33.3)	3
	5000–9999	–	4 (50)	1 (12.5)	1 (12.5)	2 (25)	8
	10,000–14,999	–	2 (28.6)	2 (28.6)	2 (28.6)	1 (14.3)	7
	15,000–24,999	–	11 (42.3)	3 (11.5)	2 (7.7)	10 (38.5)	26
	25,000–74,999	–	15 (37.5)	8 (20)	10 (25)	7 (17.5)	40
	75,000+	–	–	–	–	–	–
	Total	–	33 (39.3)	14 (16.7)	16 (19)	21 (25)	84
Total	0–4999	–	1 (10)	–	3 (30)	6 (60)	10
	5000–9999	–	4 (20)	3 (15)	2 (10)	11 (55)	20
	10,000–14,999	–	8 (23.5)	9 (26.5)	5 (14.7)	12 (35.3)	34
	15,000–24,999	–	28 (23.5)	17 (14.3)	17 (14.3)	57 (47.9)	119
	25,000–74,999	1 (0.2)	77 (16.3)	77 (16.3)	111 (23.6)	205 (43.5)	471
	75,000+	–	1 (9.1)	2 (18.2)	5 (45.5)	3 (27.3)	11
	Total	1 (0.2)	119 (17.9)	108 (16.2)	143 (21.5)	294 (44.2)	665

TABLE 26 Specimen cellularity group compared with dyskaryotic group, split by dilution level: TP

TP unmixed dilutions	Specimen cellularity group	Dyskaryotic group					Total
		0	1–	25–	50–	100–	
D1	0–2499	–	–	1 (100)	–	–	1
	2500–4999	–	–	–	1 (100)	–	1
	5000–7499	–	–	1 (100)	–	–	1
	7500–9999	–	1 (33.3)	1 (33.3)	–	1 (33.3)	3
	10,000–14,999	–	–	1 (50)	1 (50)	–	2
	15,000–24,999	–	2 (20)	2 (20)	3 (30)	3 (30)	10
	25,000–74,999	–	21 (44.7)	8 (17)	11 (23.4)	7 (14.9)	47
	75,000+	–	10 (52.6)	5 (26.3)	2 (10.5)	2 (10.5)	19
	Total	–	34 (40.5)	19 (22.6)	18 (21.4)	13 (15.5)	84
D2	0–2499	–	–	–	–	–	–
	2500–4999	–	2 (50)	1 (25)	–	1 (25)	4
	5000–7499	–	1 (50)	1 (50)	–	–	2
	7500–9999	–	2 (22.2)	2 (22.2)	3 (33.3)	2 (22.2)	9
	10,000–14,999	1 (20)	–	3 (60)	1 (20)	–	5
	15,000–24,999	–	7 (36.8)	4 (21.1)	2 (10.5)	6 (31.6)	19
	25,000–74,999	–	10 (38.5)	9 (34.6)	2 (7.7)	5 (19.2)	26
	75,000+	–	2 (50)	1 (25)	1 (25)	–	4
	Total	1 (1.4)	24 (34.8)	21 (30.4)	9 (13)	14 (20.3)	69
D3	0–2499	–	–	–	–	–	–
	2500–4999	–	2 (50)	1 (25)	1 (25)	–	4
	5000–7499	–	2 (40)	1 (20)	2 (40)	–	5
	7500–9999	–	3 (50)	–	–	3 (50)	6
	10,000–14,999	–	3 (30)	3 (30)	2 (20)	2 (20)	10
	15,000–24,999	–	7 (50)	4 (28.6)	2 (14.3)	1 (7.1)	14
	25,000–74,999	–	10 (37)	9 (33.3)	5 (18.5)	3 (11.1)	27
	75,000+	–	3 (75)	–	1 (25)	–	4
	Total	–	30 (42.9)	18 (25.7)	13 (18.6)	9 (12.9)	70
D4	0–2499	–	–	–	–	–	–
	2500–4999	–	4 (80)	–	1 (20)	–	5
	5000–7499	–	3 (50)	1 (16.7)	2 (33.3)	–	6
	7500–9999	–	1 (25)	1 (25)	–	2 (50)	4
	10,000–14,999	–	3 (42.9)	3 (42.9)	1 (14.3)	–	7
	15,000–24,999	–	5 (29.4)	3 (17.6)	3 (17.6)	6 (35.3)	17
	25,000–74,999	–	8 (38.1)	4 (19)	5 (23.8)	4 (19)	21
	75,000+	1 (33.3)	2 (66.7)	–	–	–	3
	Total	1 (1.6)	26 (41.3)	12 (19)	12 (19)	12 (19)	63

continued

TABLE 26 Specimen cellularity group compared with dyskaryotic group, split by dilution level: TP (*continued*)

TP unmixed dilutions	Specimen cellularity group	Dyskaryotic group					Total
		0	1–	25–	50–	100–	
D5	0–2499	–	1 (100)	–	–	–	1
	2500–4999	–	2 (66.7)	–	1 (33.3)	–	3
	5000–7499	–	–	–	2 (50)	2 (50)	4
	7500–9999	–	2 (50)	1 (25)	–	1 (25)	4
	10,000–14,999	–	4 (50)	2 (25)	1 (12.5)	1 (12.5)	8
	15,000–24,9999	–	6 (27.3)	6 (27.3)	6 (27.3)	4 (18.2)	22
	25,000–74,999	–	14 (63.6)	3 (13.6)	4 (18.2)	1 (4.5)	22
	75,000+	–	–	–	–	1 (100)	1
	Total	–	29 (44.6)	12 (18.5)	14 (21.5)	10 (15.4)	65
D6	0–2499	–	1 (100)	–	–	–	1
	2500–4999	–	5 (71.4)	1 (14.3)	–	1 (14.3)	7
	5000–7499	–	4 (66.7)	–	1 (16.7)	1 (16.7)	6
	7500–9999	–	2 (50)	–	1 (25)	1 (25)	4
	10,000–14,999	–	2 (28.6)	4 (57.1)	1 (14.3)	–	7
	15,000–24,999	–	10 (50)	5 (25)	3 (15)	2 (10)	20
	25,000–74,999	–	11 (50)	7 (31.8)	1 (4.5)	3 (13.6)	22
	75,000+	–	–	–	1 (100)	–	1
	Total	–	35 (51.5)	17 (25)	8 (11.8)	8 (11.8)	68
D7	0–2499	–	4 (100)	–	–	–	4
	2500–4999	–	2 (33.3)	2 (33.3)	1 (16.7)	1 (16.7)	6
	5000–7499	–	3 (50)	1 (16.7)	–	2 (33.3)	6
	7500–9999	–	2 (28.6)	3 (42.9)	1 (14.3)	1 (14.3)	7
	10,000–14,999	–	4 (36.4)	2 (18.2)	2 (18.2)	3 (27.3)	11
	15,000–24,999	–	9 (50)	4 (22.2)	3 (16.7)	2 (11.1)	18
	25,000–74,999	–	7 (70)	1 (10)	1 (10)	1 (10)	10
	75,000+	–	1 (50)	–	1 (50)	–	2
	Total	–	32 (50)	13 (20.3)	9 (14.1)	10 (15.6)	64
D8	0–2499	–	3 (75)	1 (25)	–	–	4
	2500–4999	–	3 (50)	2 (33.3)	–	1 (16.7)	6
	5000–7499	–	3 (60)	–	–	2 (40)	5
	7500–9999	–	7 (87.5)	1 (12.5)	–	–	8
	10,000–14,999	–	9 (75)	3 (25)	–	–	12
	15,000–24,999	–	11 (64.7)	4 (23.5)	1 (5.9)	1 (5.9)	17
	25,000–74,999	–	6 (66.7)	1 (11.1)	1 (11.1)	1 (11.1)	9
	75,000+	–	–	–	–	–	–
	Total	–	42 (68.9)	12 (19.7)	2 (3.3)	5 (8.2)	61

TABLE 26 Specimen cellularity group compared with dyskaryotic group, split by dilution level: TP (*continued*)

TP unmixed dilutions	Specimen cellularity group	Dyskaryotic group					Total
		0	1–	25–	50–	100–	
Total	0–2499	–	9 (81.8)	2 (18.2)	–	–	11
	2500–4999	–	20 (55.6)	7 (19.4)	5 (13.9)	4 (11.1)	36
	5000–7499	–	16 (45.7)	5 (14.3)	7 (20)	7 (20)	35
	7500–9999	–	20 (44.4)	9 (20)	5 (11.1)	11 (24.4)	45
	10,000–14,999	1 (1.6)	25 (40.3)	21 (33.9)	9 (14.5)	6 (9.7)	62
	15,000–24,999	–	57 (41.6)	32 (23.4)	23 (16.8)	25 (18.2)	137
	25,000–74,999	–	87 (47.3)	42 (22.8)	30 (16.3)	25 (13.6)	184
	75,000+	1 (2.9)	18 (52.9)	6 (17.6)	6 (17.6)	3 (8.8)	34
	Total	2 (0.4)	252 (46.3)	124 (22.8)	85 (15.6)	81 (14.9)	544

Appendix 3 Health technology assessment adequacy study: cell counting methodology for ThinPrep liquid-based cytology preparations

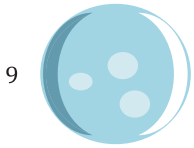
HTA Adequacy Study – cell counting methodology for Thin Prep LBC preparations

Please read this sheet before you start the cell counts. It is important to the study that all participants conduct the cell counts in the same way to allow a standardised approach and comparison of data.
The details of each count must be entered on the electronic Cell Counting Spreadsheet provided. The database will ask for the FN value of your eyepieces (typically FN 22, 20 or 16); whether or not your microscope produces a 'true' or an inverted image; and the quadrant of the deposit you have selected to perform the counts (with slide label to the left). The ten counts are then recorded individually. Please follow these steps:

- Examine the slide naked eye. The deposit may have a denser peripheral rim and paler centre which may contain holes (Figure 1). Sometimes the deposit may be uneven with part of the edge appearing darker or lighter (Figure 2). Choose a quadrant (12, 3, 6 or 9 o'clock) which is **neither hypo- nor hypercellular**. In Figure 2 the 9 o'clock and 3 o'clock positions should therefore be avoided.

Figure 1

Figure 2



- The high power fields (x40 objective) used for counting cannot be preselected. Start at the edge of one quadrant of the deposit and work in towards the centre of the deposit. You may therefore be starting your count at the 12 o'clock, 3 o'clock, 6 o'clock or 9 o'clock positions and be counting in either a vertical or horizontal direction.
- Counts should be performed on 10 fields working from the edge of the deposit towards the centre but missing out every alternate field. Do not introduce gaps between fields. Do not pass over a field if it is either particularly hypo- or hypercellular. If there is no cellular material in the field record the result as zero. Use non-cellular debris or approximate travel on the stage controls to gauge the next field if there is no cellular material in the field.
- Only squamous cells with nuclei are counted but these can be of mature or parabasal / metaplastic type. Both single cells and cells in groups must be counted. Note that very pale nuclei if still visible are counted. Free nuclei are not counted. Anucleate squames / fragments of squamous cytoplasm are not counted. Syncytial aggregates of squamous cells as seen in cytolysis can be counted according to the number of nuclei they contain even if the cytoplasmic margins of individual cells are not identifiable.
- Cells at the edge of the field are counted if the entire circumference of the nucleus can be seen. If only part of the nucleus is visible do not count. Do not move the field to see cells at the edge.
- Counting must include cells on all planes of focus. When there are exceptionally thick groups of cells, which cannot be counted individually, an estimate of cellularity can be used. A full quadrant of a high power field contains approximately 1,000 small parabasals and approximately 750 mature squames. This figure can be scaled up or down to match the amount of the field covered e.g. a sheet of small parabasal squamous cells covering half of the field would equate to approximately 2,000 cells. **Please endeavour to count the cells. The default value should only be used on very rare occasions.**

Appendix 4 Health technology assessment adequacy study: cell counting methodology for SurePath liquid-based cytology preparations

HTA Adequacy Study – cell counting methodology for SurePath LBC preparations

Please read this sheet before you start the cell counts. It is important to the study that all participants conduct the cell counts in the same way to allow a standardised approach and comparison of data.

The details of each count must be entered on the electronic Cell Counting Spreadsheet provided. The spreadsheet will ask for the FN value of your eyepieces (typically FN 22, 20 or 16); whether or not your microscope produces a 'true' -or an inverted image; and the quadrant of the deposit you have selected to perform the counts (with slide label to the left). The ten counts are then recorded individually. Please follow these steps:

- Examine the slide naked eye. The deposit will often look completely even (Figure 1), but sometimes a paler arc may be noted at the edge (Figure 2). Very rarely part of the circumference of the deposit may look slightly darker. Choose a quadrant (12, 3, 6 or 9 o'clock) which is **neither hypo- nor hypercellular**. In Figure 2 the 3 o'clock position should therefore be avoided.

Figure 1

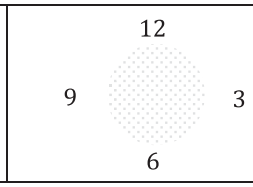


Figure 2

- The high power fields (x40 objective) used for counting cannot be preselected. Start at the edge of one quadrant of the deposit and work in towards the centre of the deposit. You may therefore be starting your count at the 12 o'clock, 3 o'clock, 6 o'c1ock or 9 o'clock positions and be counting in either a vertical or horizontal direction.

- Counts should be performed on 10 consecutive fields, working from the edge of the deposit towards the centre. Do not allow fields to overlap, but equally do not introduce gaps between fields. Do not pass over a field if it is either particularly hypo- or hypercellular. If a field is hypercellular, divide it into halves or quarters and try to remember as you go along what you have already counted.

- Only squamous cells with nuclei are counted but these can be of mature or parabasal / metaplastic type. Both single cells and cells in groups must be counted. Note that very pale nuclei if still visible are counted. Free nuclei are not counted. Anucleate squames / fragments of squamous cytoplasm are not counted. Syncytial aggregates of squamous cells as seen in cytolysis can be counted according to the number of nuclei they contain even if the cytoplasmic margins of individual cells are not identifiable.

- Cells at the edge of the field are counted if the entire circumference of the nucleus can be seen. If only part of the nucleus is visible do not count. Do not move the field to see cells at the edge.

- Counting must include cells on all planes of focus. When there are exceptionally thick groups of cells, which cannot be counted individually, an estimate of cellularity can be used. A full quadrant of a high power field contains approximately 1,000 small parabasals and approximately 750 mature squames. This figure can be scaled up or down to match the amount of the field covered e.g. a sheet of small para basal squamous cells covering half of the field would equate to approximately 2,000 cells.

Appendix 5 Protocol for counting dyskaryotic cells in ThinPrep cervical samples

Protocol for Counting Dyskaryotic cells in ThinPrep cervical samples

Basic principles

- Only well preserved and clearly dyskaryotic cells should be included in the count. Cells which are deemed equivocal should be omitted. Degenerate cells and dyskeratotic (but NOT clearly dyskaryotic) cells even when they lie adjacent to unequivocally dyskaryotic cells should not be included.
- Care should be taken to avoid over-calling reactive change and, in particular, reactive change in metaplastic squames and in endocervical cells as dyskaryotic.
- Only dyskaryotic cells in which the entire nuclear circumference is within the optical field should be counted. As with previous counting guidance, dyskaryotic cells which lie on the edge of a field and which are transected by it should not be counted.
- An estimate of the numbers of dyskaryotic cells in HCGs should also follow previous counting guidance i.e. a full quadrant of a high power field contains approximately 1,000 small parabasals and approximately 750 mature squames. This figure can be scaled up or down to match the amount of the field covered e.g. a sheet of small parabasal squamous cells covering half of the field would equate to approximately 2,000 cells.

Counting methodology

- 40 high power fields (X40 objective) should be counted in accordance with the illustration in Figure 1 and the counts entered directly into the Excel spreadsheet provided. The type and presentation of dyskaryotic cells including the nuclear changes and presence/absence of koilocytosis should also be recorded using the drop-down menus in the spreadsheet which follow the cell counts.
- The operator should commence at the edge of the deposit at the 9 o'clock position and should perform 10 counts of alternating and non-overlapping fields working towards the centre of the deposit.
- The operator should then move vertically upwards (i.e. towards 12 o'clock position) by 1 high power field and move to the edge of the slide at the 3 o'clock position. 10 counts on alternate and non-overlapping fields should be performed working from the edge to the centre of the deposit.
- The operator should then move to the edge of the slide at the 11 o'clock position and working vertically downwards should perform 5 counts on alternate and non-overlapping fields.
- The operator should then move to the edge of the slide at the 7 o'clock position and working vertically upwards should perform 5 counts on alternate and non-overlapping fields.
- The operator should then move to the edge of the slide at the 1 o'clock position and working vertically downwards should perform 5 counts on alternate and non-overlapping fields.

- The operator should then move to the edge of the slide at the 5 o'clock position and working vertically upwards should perform 5 counts on alternate and non-overlapping fields.

If the above methodology which counts a total of 40 fields fails to detect any dyskaryotic cells then the operator should perform a full manual screen avoiding any overlap of fields and counting all dyskaryotic cells which are detected. The total count should then be entered as a single entry on the spreadsheet in the column 'Total number dyskaryotic cells'.

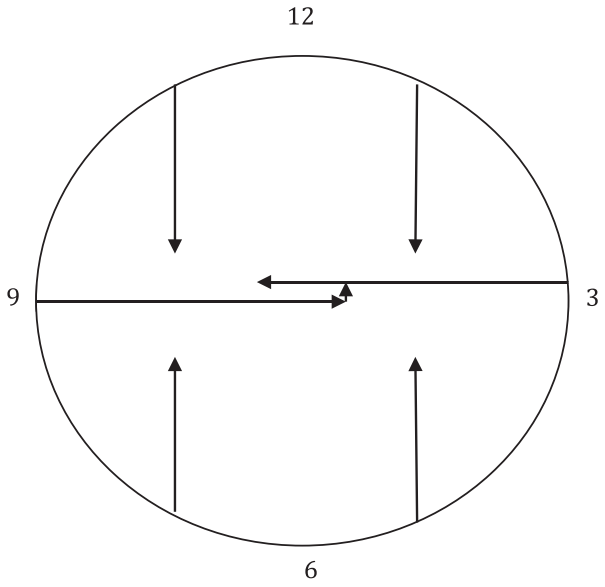


Figure 1- Diagrammatic representation of counting pattern

Dr LS Turnbull
May 2010

Appendix 6 Preparation of slides for the dilution studies

Preparation of slides for the dilution studies

In this part of the study the impact of total cellularity and the relative proportion of abnormal cells on the screener's ability to correctly detect that abnormality is being examined. In the protocols we have committed to making a further 8 or 9 slides from selected positive cases and either diluting those samples (unmixed dilutions) or mixing them with varying amounts of a negative sample (mixed dilutions). The protocol reads as follows:

'A total of 180 SurePath™ and 180 Thin Prep® cases will be selected from material routinely accessioned at the Royal Liverpool University Hospital and the Manchester Cytology Centre, which display a range of histologically confirmed low and high grade cytological abnormalities. Following selection, the cases will be entered on the study database and anonymised to ensure compliance with the Human Tissue Act 2004. The cases will range from those containing plentiful hyperchromatic (darkly staining) dyskaryotic cells, to those which have scanty abnormal cells; are pale staining; show minimal nuclear changes or form microbiopsy fragments. The latter sub-types are known to cause diagnostic problems.

Eight preparations will be made from each sample. Serial dilutions with approximate cellularities of 5-10k, 10-15k, 15-20k, 20-25k, 25-35k, 35-45k, 45-55k and 55+k will be made from half of the cases. The range of dilutions is skewed towards preparations of lower cellularity as these are expected to have higher false negative rates. The remaining cases will be mixed with known negative cases in varying proportion to establish sets of slides containing <25, 25-49, 50-99, 100-149, 150-199, 200-399, 400-799, 800-1600 and 1600+abnormal cells. In total 3,000 new slides will be prepared. '

The methodology will have to be different for the two systems as the means by which the systems acquire cells is different. A pilot of 5 SurePath and 5 ThinPrep cases for the straight dilutional component of this part of the study should be performed and assessed fully before processing others.

The cellularity of the original slide can be determined by a cell count. Theoretically, at least for Surepath, we could tailor the dilutions to suit each sample. This would be unwieldy and could be very complicated for the technical staff. It is suggested therefore that a standard methodology is used for all cases accepting that there may be a greater spread of cellularities than is quoted in the original protocol. Assuming an average total cellularity of 60k (original slide) then the dilutions below would give approximate cellularities of 45k, 30k, 25k, 22.5k, 20k, 15k, 10k, 5k and 2.5k. The original slide together with the dilutions would provide a total of 10 slides for each case. Following the cell counts 8 slides, which best suit the required cellularities, will be selected for morphological assessment by the participating laboratories.

SurePath

- The residual samples are stored in tubes and have been 'topped up' with 2ml of collection fluid.

- The tube should be recentrifuged, drained and the cell deposit topped up with 1ml of collection fluid to produce a working solution. (The working solution will be diluted with collection fluid to produce specimens of varying cellularity.)
- The slides to be used in the PrepStain machine must be appropriately labelled with the case accession number followed by the dilution number to ensure traceability. (If the methodology is successful it may be easier in subsequent runs to complete the same dilution for a block of cases rather than having differing dilutions on a single run).
- The PrepStain machine will be used as normal but there will be no samples in the sample test tube rack; rather the sample will be pipetted directly into the sedimentation chamber. The SurePath system requires that a total 200µl of fluid is placed in the sedimentation chamber. For this part of the study, the 200 µl will comprise a combination of sample and collection fluid. The volumes for each dilution are given in Table 1 below. It is important that the collection fluid is added before the sample to avoid the latter drying out.
- Processing of the samples is then completed as per usual.

Table 1 – Schedule of dilutions for SurePath

Slide label	Expected Sample cellularity	Volume of sample to be added (µl)	Volume of diluting collection fluid (µl)
Accession no – D1	Original slide		
Accession no – D2	45k	150	50
Accession no – D3	30k	100	100
Accession no – D4	25k	83	117
Accession no – D5	22.5K	75	125
Accession no – D6	20k	67	133
Accession no – D7	15k	50	150
Accession no – D8	10k	33	167
Accession no – D9	5k	17	183
Accession no – D10	2.5K	9	191

ThinPrep

The production of dilutional slides for ThinPrep is less easy than for SurePath and will almost certainly require more trial and error. In order to produce poorly cellular slides the system will need to be offered super-dilute samples and the processor allowed to filter all or most of their volume. This effectively over-rides the error message which the system will display.

- The residual sample is currently stored in the original vial.

- Only cases where >15ml sample remains in the vial can be selected. A total of 12.945ml of sample is required to produce the suggested dilutions.
- The slides to be used in the T2000 machine must be appropriately labelled with the case accession number followed by the dilution number to ensure traceability. (If the methodology is successful it may be easier in subsequent runs to complete the same dilution for a block of cases rather than having differing dilutions on a single run).
- The residual sample is divided between 9 vials in the volumes given in Table 2 and topped up with PreservCyte to a total volume of 20ml. The vials can be recycled vials as minor contamination is not an issue.
- The vials are then offered to the T2000 in the usual way and stained/coverslipped prior to counting.

The following schedule (Table 2) is suggested as a trial:

Table 2 – Schedule of dilutions for ThinPrep

Slide label		Volume of sample to be added (ml)	Volume of PreservCyte (ml)
Accession no – D1	Original slide		
Accession no – D2	45k	5.0	15.0
Accession no – D3	30k	3.0	17.0
Accession no – D4	25k	1.5	18.5
Accession no – D5	22.5K	1.0	19.0
Accession no – D6	20k	0.75	19.25
Accession no – D7	15k	0.5	19.5
Accession no – D8	10k	0.25	19.75
Accession no – D9	5k	0.13	19.87
Accession no – D10	2.5K	0.07	19.93

Varying the proportion of normal/abnormal cells

This arm of the protocol requires that abnormal cases are mixed with known negatives in differing quantities to vary the number and relative proportion of abnormal cells. The protocols for these preparations again differ between the two LBC systems but for ease of production I would suggest that within LBC system all samples are handled identically.

At the recent Management Group meeting it was decided that to facilitate the circulation, 88 rather than 90 cases would be used and each of these would have 8 slides. The first slide will be the original preparation, hence an additional 7 slides will be produced.

SurePath

- 90 cases have already been earmarked for this arm of the study and the test tubes which have been topped up with 2ml of collection fluid are being held in reserve in the laboratory. Use numbers 1-88 only. Hold the other 2 cases in reserve in case there are problems. To make the preparations the residual material from 88 'negative' cases will also be needed. The original slides from these negative cases should be of average cellularity with no particularly recognisable background pattern i.e. not atrophic or especially inflammatory. LT will select these cases from current workload.
- Two 'stock' solutions should be made up intially.
- For 'stock 1'. Vortex test tube from case to homogenise sample. Pipette 500 µl of sample into clean test tube and add 4.5 ml of collection fliud. Total volume in test tube will then be 5.0ml.
- For 'stock 2'. Vortex test tube from case to homogenise sample. Pipette 50 µl of sample into clean test tube and add 5.0 ml of collection fliud. Total volume in test tube will then be 5.05ml.
- The slides to be used in the PrepStain machine must be appropriately labelled with the case accession number followed by the dilution number to ensure traceability.
- The PrepStain machine will be used as normal but there will be no samples in the sample test tube rack; rather the sample will be pipetted directly into the sedimentation chamber. The SurePath system requires that a total 200µl of fluid is placed in the sedimentation chamber. For this part of the study, the 200 µl will comprise a combination of 'stock' sample and a negative case. The volumes for each dilution are given in Table 1 below.
- Processing of the samples is then completed as per usual.

D1 Original slide		
	Negative sample vol in µl	Test sample vol in µl
		USE STOCK 1
D2	100	100
D3	150	50
D4	175	25
D5	190	10
		USE STOCK 2

D6	150	50
D7	175	25
D8	190	10

ThinPrep

- 90 cases have already been earmarked for this arm of the study and the vials are being held in reserve in the laboratory. Use numbers 1-88 only. Hold the other 2 cases in reserve in case there are problems. To make the preparations the residual material from 88 'negative' cases will also be needed. The original slides from these negative cases should be of average cellularity with no particularly recognisable background pattern (i.e. not atrophic or especially inflammatory) and each will require a minimum residual volume in the vial of at least 12ml and preferrably 15ml.

- The test vial should be centrifuged to obtain a cell button. The cell button should then be resuspended in 3ml of PreservCyte.

- The slides to be used in the T2000 machine must be appropriately labelled with the case accession number followed by the dilution number to ensure traceability.

- Decreasing volumes of test sample are then added to the residual volume in known negative cases as per following table:

D1 Original slide	
	Test sample vol in µl
D2	1,000
D3	500
D4	250
D5	100
D6	50
D7	25
D8	10

- The vials are then offered to the T2000 in the usual way and stained/coverslipped prior to counting.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library