<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td colspan="6">Bowen et al (2014) USA</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

January 2009 – December 2010</td><td>To determine the rate of revised diagnosis and subsequent impact on therapy following a second review</td><td>N=1010

N=683 (67.6%) mandatory reviews
N=327 (32.4%) outside consultations

N=142 (14%) referred from academic centres
N=868 (86%) referred from non-academic centres

Exclusions
Myeloid neoplasms
Acute lymphoblastic leukaemia
Plasma cell myeloma
Staging bone marrows for non-haematological malignancies
Cases sent without a primary diagnosis

Inclusions
Lymph Nodes and extranodal tissues that were reactive or benign</td><td>Second Review Diagnosis</td><td>Primary referral diagnosis</td><td>Diagnostic Discrepancies

• Second review resulted in no change to diagnosis in 83% of cases
• In 17% of cases second review resulted in a changed or modified diagnosis
  ○ 14.8% were considered major discrepancies and 12.9% resulted in significant changes to therapy
  ○ 2.2% were considered minor discrepancies and so were grouped with the agreement cases
• Overall agreement was 85.2% when considering only major discrepancies
• The largest category of discrepant cases was one in which diagnosis was revised from one type of lymphoma to another (6.5%) with change from one type of B-NHL to another B-NHL being the most common revision within this group (4.3%)
• 3% of grading discrepancies occurred in Follicular Lymphoma with most diagnoses changing from low grade to high grade on second review
• 2.8% of discrepancies occurred in benign entities originally diagnosed as lymphoma or vice versa.
• Imprecise or unclear diagnoses occurred in 2.1% of discordant cases
• There was a significantly higher rate of discordance in diagnoses from non-academic centres compared with academic centres (15.8% versus 8.5%, p=0.022)
• There were similar rates of discordance between referral cases and consultation cases (15% versus 13.5%, p=0.42)
• Excision biopsies (61.9%) had a significantly higher rate of discordance compared to other biopsy types (needle core, punch biopsy, shave biopsy) (17.9% versus 9.6%, p=0.0003)
• Biopsy site (lymph nodes (52.1%), bone marrow (14.3%), soft tissue (8.5%), gastrointestinal tract (6.3%), skin (5.8%)) was not a significant factor affecting disagreement rate (p=0.20).
• Cases requiring additional investigative studies (51.5%) had a significantly higher rate of revised diagnosis compared to ones not requiring additional studies (20.6% versus 8.6%, p<0.0001).

Comments
Cases were divided into two groups – ‘mandatory reviews’ and ‘outside consultations’

Mandatory Reviews: patient referral to NLSG for clinical management

Outside Consultation: pathology slides and materials sent for a second opinion

Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported – likely consecutive

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (Haematology patients)

Low Risk of Bias</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>
<table>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>No

High Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>No

Unclear Risk of Bias</td></tr>
</table>
</td></tr>
<tr><td colspan="6">Chang et al (2014)</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

2003-2011</td><td>To review the final diagnoses made by general pathologists and analyse the discrepancies between referral and review diagnosis</td><td>N=395 (406 specimens)

Cases transferred for treatment or for second opinion were excluded</td><td>Expert Review</td><td>Initial Diagnosis</td><td>Diagnostic Discrepancies

• Turnaround time was 2.3 days (0-19 days)
• 95% of cases sent for review were haematological cases and 5% were non-haematological lesions
• Pathology review resulted in major revisions in 55% of cases, minor revisions in 5% of cases and insignificant revision or agreement in 40% of cases
• The major discrepancy category (52%) was the most common group consisted of ambiguous and non-diagnostic reports and the more common lymphoma types were diffuse large B cell lymphoma, marginal zone lymphoma and follicular lymphoma
• In Group 2, the revision of lymphoma typing (23%), the most common entities were diffuse large B cell lymphoma, Hodgkin Lymphoma and plasmacytoma/myeloma
• Group 3 represented cases from malignant to benign diagnosis (n=32, 14.4%)
• Group 4 was the easily missed lymphomas (4%), group 5 consisted of haematologic tumours revised as non-haematologic tumours (5%) and group 6 was non-lymphoma tumours revised as lymphomas (1%)
• Review diagnosis results in 259 cases of lymphoma (72% B-cell and Hodgkin lymphoma, 28% T/natural killer cell lymphomas)
• Comparison between referral and review diagnosis showed a lymphoma concordance rate of 39% (101/259) in total, 41% (77/187) for B cell lymphoma and 33% (24/72) for T/NK cell lymphomas respectively.

Comment
Major discrepancies – those that would alter management

Minor discrepancies – those that would not fundamentally alter management although a different diagnosis was given

Non-diagnostic reports – diagnosis not given by referral diagnosis

Ambiguous original reports – diagnosis not sufficiently specific to generate a treatment plan</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>

Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported – likely consecutive

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (Lymphoma patients)

Low Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>No

High Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Unclear Risk of Bias</td></tr>
</table>

</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td colspan="6">Engel Nitz et al (2014) USA</td></tr>
<tr>
<td>Retrospective Study

Laboratory Setting

July 2005 – June 2011</td>
<td>To compare diagnostic changes, patterns of additional testing, treatment decisions and health care costs for patients with suspected haematological malignancies/conditions whose diagnostic tests were managed by specialty haematology laboratories and other commercial laboratories.</td>
<td>Initial laboratory population N=34,904 – patients with non-haematological cancer or any other non-haematological condition on bone marrow biopsy claims were excluded from analysis.

N=24,664 patients

Genoptix N=1,387
Large Labs N=4,162
Other Controls (community hospital labs) N=19,115

Academic labs that sponsor haematopathology fellowships were excluded due to the likelihood of a higher percentage of referral cases.

Patients with suspected haematological malignancies/conditions who had a bone marrow procedure (biopsy/aspirate) INDEX DATE

Patients were grouped according to diagnosis – Myelodysplastic Syndrome, myeloproliferative neoplasm, Chronic</td>
<td>Initial interim diagnosis (based on date of first non-laboratory claim with a diagnosis of haematological malignancy/disease in the primary position at least 3 days after and <1 year post-index date

Laboratory tests in the 30 days post biopsy were identified</td>
<td>Final Diagnosis</td>
<td>

Diagnostic Uncertainty following initial diagnostic uncertainty (using 2 definitions comparing haematological diagnosis between initial interim and final diagnoses)
Stability of Diagnosis (at least 1 haematological condition that was the same between the two time points, excluding disease progression or haematological signs/symptoms)
Number of tests performed
Repeat bone marrow studies
Time to final diagnosis
Changes in chemotherapy in the 60 days post-biopsy
Testing Costs
All cause health care costs

Baseline Characteristics

Patients in other laboratories were younger compared with Genoptix and Large lab patients (p<0.001) and were more likely to be enrolled in Medicare advantage plans (p<0.001)
Genoptix patients were more likely to be located in the south
Patients in the ‘other laboratory’ cohort were more likely to have had chemotherapy or radiotherapy.

Diagnostic Characteristics

Patients in the Genoptix cohort were more likely to undergo more complex diagnostic testing during the initial 30 day testing period.
Patients in the other lab cohort were less likely to undergo complex diagnostic testing and when done, these tests were more likely to be performed at a different lab type.

<table>
<tr><th></th><th>Cytogenetics/FISH</th><th>Molecular Diagnostics</th></tr>
<tr><td>Genoptix</td><td>95.96%</td><td>26.03%</td></tr>
<tr><td>Large laboratory</td><td>80.78%</td><td>14.27%</td></tr>
<tr><td>Other laboratory</td><td>51.68%</td><td>9.31%</td></tr>
</table>

The number of tests varied across the 1 year follow-up period though the majority of patients received 1 bone marrow biopsy
The large lab cohort had the fewest total test and average time to final diagnosis ranged from 36 days for Genoptix to 41 days for the other lab cohort.
Median time to final diagnosis was roughly 2 weeks. The Cox proportional hazard ratio of reaching a final diagnosis by any point in the initial 30 days testing period were 1.002 (p=0.0029) for the Genoptix Cohort and 0.95 (p=0.0002) for the large lab cohort (other lab cohort as the reference group).
At any point in the 30 day testing period, the Genoptix cohort had a 23% higher hazard of having reached a final diagnosis compared with the other lab cohort (HR=1.23, p=0.0007) and the large lab cohort had a 10% higher hazard (HR=1.10, p=0.005) compared with the other lab cohort.
Fewer patients in the Genoptix cohort underwent repeat marrow biopsies, with difference remaining after adjustment for type of haematological malignancy and other characteristics

</td>
</tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td>Lymphoid Leukaemia, non-Hodgkin lymphoma, multiple myeloma, other haematological cancer, non-cancer haematological condition</td><td></td><td></td><td>
<table>
<tr><th></th><th>Repeat Marrow Biopsy</th><th>Odds Ratio</th><th>P value</th></tr>
<tr><td>Genoptix</td><td>9.59%</td><td>0.307 (0.255-0.371)</td><td>P<0.001</td></tr>
<tr><td>Large laboratory</td><td>17.11%</td><td>0.563 (0.514-0.617)</td><td>P<0.001</td></tr>
<tr><td>Other laboratory</td><td>28.16%</td><td colspan="2">Reference Group</td></tr>
</table>
Stability of initial diagnosis varied across the cohorts
<table>
<tr><th></th><th>Unstable Diagnoses</th><th>Odds Ratio</th><th>P value</th></tr>
<tr><td>Genoptix</td><td>6.16%</td><td>0.87 (0.68-1.10)</td><td>0.2427</td></tr>
<tr><td>Large laboratory</td><td>8.04%</td><td>0.99 (0.87-1.13)</td><td>0.9014</td></tr>
<tr><td>Other laboratory</td><td>9.73%</td><td colspan="2">Reference Group</td></tr>
</table>
The percentage of diagnoses changes was lower in the Geneoptix cohort
<table>
<tr><th></th><th>Change in Diagnosis</th><th>Odds Ratio</th><th>P Value</th></tr>
<tr><td>Genoptix</td><td>7.88%</td><td>0.82 (0.72-0.94)</td><td>0.004</td></tr>
<tr><td>Large laboratory</td><td>11.19%</td><td>0.94 (0.87-1.02)</td><td>0.1256</td></tr>
<tr><td>Other laboratory</td><td>14.08%</td><td colspan="2">Reference Group</td></tr>
</table>
Comments
Length of follow up: First appearance was followed up for 1 year post index date

Genoptix - a specialised haematopathology lab which designed a specific diagnostic workflow to address the main concerns associated with diagnostic testing in the community oncology setting (tests ordered, sampling error, and interpretation/integration errors).
</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported – likely consecutive

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (Haematology patients)

Low Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>Unclear - there was not enough information reported in the study to determine whether the comparisons were those outlined in the PICO. Personal communication from the author provided more detail which suggested that the comparisons were more closely matched to those of interest than was first though, however some of the additional information provided also suggested there were some differences between the comparisons which meant that this study did not completely address the PICO.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>No

High Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Unclear Risk of Bias</td></tr>
</table></td></tr>
<tr><td colspan="6">Gundlapalli et al (2009) USA</td></tr>
<tr><td>Survey</td><td>To address the hypotheses that clinical providers perceive composite laboratory reports to be important for the care of complex</td><td>N=10 clinical staff

Clinical staff involved in the Myeloma program and who routinely accessed the patient labs</td><td>Survery and interview</td><td>None</td><td>End User Survey
• Team members spent an average of 18 minutes per patient gathering lab data and an average of 4 minutes per patients on protein immunology labs.
• 6/10 responders reported being familiar with or having used the ‘trend’ or ‘graph’ feature of the EMR to view serial labs with numeric results
• All providers reported accessing free text reports of serum protein electrophoresis and immune fixation electrophoresis because it was the only way to identify the presence of a myeloma protein, its type and quantitation.</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td>patients and that such reports can be generated using laboratory informatics methods</td><td>Average experience was 9 years (range 1-30 years)

All accessed the electronic medical record multiple times per day with the laboratory results screen the most accessed tab.</td><td></td><td></td><td>

7/10 reported accessing and viewing pdf files of actual gels
All 10 reported they would be in favour of a single report with the ability to view serial changes in key myeloma biomarkers
8/10 were willing to collaborate with informatics teams to work up an ideal composite report and were willing to participate in a validation study.
All 10 supported the idea of providing a composite report directly to the patient.
The primary elements identified were that access to and downloading of disparate protein immunology lab data and free text interpretations were challenging and time consuming and the provision of a composite report would be beneficial to patient care and improve work flow.

Data Flow of Laboratory Orders and Results

During 2007, a total of 4699 protein immunology tests were performed on 1450 unique patients, these tests are performed multiple times on accessing and correlation of even the last 3 results of tests reported in free text poses a challenge

Comments:

Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Unclear – clinic staff

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>N/A</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Unclear Risk of Bias</td></tr>
</table>
</td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results
Herrera et al (2014) USA					
Retrospective Study Laboratory Setting April 2007-June 2012	To evaluate the rate of diagnostic concordance between referring centre diagnoses and expert haematology review for 4 subtypes of T-cell lymphoma	N=89 *Inclusion* Documented pathologic review at a referring centre before expert haematology review Final diagnosis of 1 of the following 4 TCL WHO subtypes; PTCL-NOS, AITL, ALK negative ALCL and ALK positive ALCL *Exclusion* Primary presentation to an NCCN centre so no referring pathology Incomplete or insufficient data for analysis	Review of primary diagnosis at an NCCN centrte	Primary diagnosis at a referring centre	*Concordance between diagnoses* • Overall concordance rate was 44% (n=57 patients with concordant results) and the discordant rate was 24% (n=32 patients with discordant results). • 32% of patients (n=42) were referred for a second opinion with additional biopsy or further work-up suggested • Rates of pathologic discordance were 19% for PTCL-NOS, 33% for AITL, 34% for ALK negative ALCL and 6% for ALK positive ALCL • Discordance rates among patients referred for a second opinion with final diagnosis were 38% for PTCL-NOS, 50% for AITL, 38% for ALK negative ALCL and 7% for ALK positive ALCL • 47% (15/32) of patients were reclassified based on a different interpretation of available data or noncontributary additional studies • 53% (17/32) of patients with discordant results had additional studies performed at the NCCN centre which led to a different diagnosis. • 86% (n=112) of patients had an excision biopsy sample submitted for review by an NCCN centre and no association was observed between biopsy type and pathologic concordance among patients referred with a final diagnosis (p=0.18) or between biopsy type and whether a final diagnosis was rendered at the referring centre (p=0.09). • Additional testing was performed at the referring centre before second opinion referral in 95% of cases (IHC stains=84%; flow cytometry=52%; TCR gene rearrangement testing=36% and FISH=6%). There was no association between pathologic concordance or discordance and the type of additional tests performed (IHC p=0.66, flow cytometry p =0.83, TCR gene rearrangement testing p = 0.5, IHC+flow cytometry p=0.825, IHC+flow cytometry+TCR testing p=0.6). • Additional testing performed in at the NCCN centre included IHC stains (53%), flow cytometry (18%), TCR gene rearrangement (18%) and FISH (6%). • Median number of IHC stains performed at the NCCN centres was 2 (range 0-29) compared with 11 (range 0-35) at the referring centres • Median duration of time spent reviewing a case at the NCCN centre was 5 days (range 1-34 days) • 72% of cases were reviewed by a single pathologist and 28% were referred for intradepartmental consulation compared with 76% and 24% at the referring centres. ***Comments*** Pathological concordance was defined as the same pathological diagnosis at both the referring centre and the NCCN centre when considering all supporting documentation including pathology reports, immunohistochemistry, flow cytometry, fluorescence in-situ hybridization and cytogenetics, T-cell gene rearrangement studies and physician progress notes Review of records was carried out by 3 of the authors to determine concordance

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>N/A</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Unclear Risk of Bias</td></tr>
</table>
</td></tr>
<tr><td colspan="6">Irving et al (2009) UK</td></tr>
<tr><td>Report

Laboratory Setting</td><td>To show that the standardised protocol has high sensitivity and technical applicability, has good concordance with the gold standard molecular based analysis and is highly reproducible between laboratories across different instrument platforms.</td><td>No details</td><td>Standardised protocol for flow cytometry</td><td>Gold standard molecular technique</td><td>Internal and External quality assurance testing of Flow minimal residual disease
• QA testing consisted of mock MRD sample posted to all 6 network laboratories for analysis and interpretation (n=15 samples prepared by laboratories within the network using fresh material and n=6 provided by the UK National External Quality Assessment Scheme using mock samples prepared with fixed, stabilised material)
• List mode data files of MRD samples acquired in one centre were analysed by all network laboratories to assess gating strategies (n=2)
• Gives a total of 23 quality assessment exercises with 42 separate LAIP analyses

• Interlaboratory correlation coefficient ranged from 0.97 to 0.99
• Interlaboratory agreement on risk category compared to the consensus risk was 100% for 4 laboratories, 90% for one lab and 80% for one lab. One discordant example was attributed to inappropriate gating which was subsequently standardised during group workshops.

Sensitivity and variability of the standardised method
• Sensitivity of the assay was assessed by spiking leukaemic blasts with a known LAIP into normal bone marrow and preparing serial dilutions down to 0.01%. a sensitivity of 0,01% was confirmed for all LAIP combinations tested (CD38, CD45, CD58 and CD66c).
• Interassay variability was assessed using mock MRD replicates analysed using 2 different cytometers. The coefficient variation ranged from 2.2%-4.1%, 3.14%-5.47% and 10.21%-13.13% for 10%, 0.5% and 0.05% MRD mocks respectively.</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>
Applicability of the standardised method in prospective samples
• 182/206 patients with diagnostic precursor B-lineage ALL had 2 or more sensitive LAIPs for an applicability of 88.3%
• 45/182 (24.7%) of patients were classified high risk at day 28.

Comparison of minimal residual disease as measured by PCR and by flow cytometry
• MRD quantification of bone marrow aspirates was performed by both PCR and flow cytometry in 134 children.
• 90 samples were low risk by both methods, 25 were high risk by both methods, 8 were high risk by flow cytometry but low risk by molecular and 11 were low risk by flow and high risk by molecular.
• Excluding the 90 cases below the threshold of both methods, the percentage of cases in which logPCR and log Flow MRD were within half a log was 47.6% and within one log was 76.2%.
• The risk category concordance was 79% at day 28 and 100% at week 11 for a combined figure of 86%
• In the 25 high risk samples, correlation was high (r=0.76).
• The majority of the discordant samples were around the threshold level and in 8 sample, MRD was detectable by both techniques but did not attain the 0.01% level in both assays.

Comments:

Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>N/A</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Yes (UK study)

Low Risk of Bias</td></tr>
</table>
</td></tr>
<tr><td colspan="6">LaCasce et al (2005) USA</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr>
<td>Retrospective Study

Laboratory Setting

July 1, 200 and December 31, 2004</td>
<td>To determine the rate of discordance for 5 common B-cell NHL diagnoses in five tertiary centres participating in a large national lymphoma database
The determine whether additional information was obtained at the National Comprehensive Cancer Network (NCCN) centre
To estimate the likely impact of a change in diagnosis on treatment</td>
<td>N=928 patients presented with newly diagnosed NHL

N=731 referred from other centres and had a documented pathologic diagnosis of one of 10 NHL subtypes before presentation at the NCCN

N=66 patients for whom the referring diagnosis and the NCCN diagnosis were discordant

Patients with newly diagnosed NHL (≤90 days from diagnostic biopsy date to first NCCN presentation)
Documented pathologic diagnosis assessed at a referral centre
Final diagnosis of follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL), Mantle cell lymphoma (MCL), small lymphatic lymphoma (SLL), nodal marginal zone lymphoma (NMZ), extranodal marginal zone lymphoma (EMZ) or splenic marginal zone lymphoma (SMZ)</td>
<td>Pathologic diagnosis from the referral centre was compared with the final WHO diagnosis at the NCCN centres

Etiology of the discordance was investigated along with the potential impact on treatment.

A random sample of concordant cases (10%) were also reviewed</td>
<td>No Details</td>
<td>
Discordance Rates
42/66 patients were considered truly discordant after central and site review and 1 additional pathologically discordant case was identified among the sample of concordant cases reviewed and was included in the analysis

• Overall pathologic discordance rate was 6% (95% 4%-8%)
• Pathologic concordance was highest for DLBCL, FL and MZL
• Final diagnosis with the highest proportion of pathologic discordance was FL3 (13%) though the total number of cases was small (=32)
• Reasons for a change in pathologic diagnosis included: preliminary diagnosis with further evaluation recommended (n=4), different interpretation of the existing data (n=19), one or more additional biopsies performed (n=9), other studies including immunoperoxidasae stains were performed (n=11).
• Treatment category discordance occurred in 5% (95% CI 3%-7%) of cases overall and in 81% (35/43) patients in whom pathology was discordant.
• 2% of patients with DLBCL were assigned a pathological diagnosis at the referral centre which resulted in less aggressive treatment thus missing a chance for cure
• All patients who with FL3 who were pathologically discordant were also treatment discordant with original diagnosis classified as indolent.
• Fine needle aspiration and core biopsy accounted for 9% (n=68) and 19% (n=142) of initial biopsies at referral sites with no statistically significant difference in concordance between those who had FNA or core biopsy or other biopsy types (94%, 93% and 94% respectively, p=0.76)
• Proportions of nodal and extra nodal referrals were 61% (n=473) and 34% (n=258) respectively and there was no statistically significant difference in concordance between nodal and extranodal referral specimen (94% versus 95%, p=0.47)
60% (n=437) of cases had ancillary testing prior to presentation at NCCN but there was no statistically significant difference in concordance between referral specimens with and without ancillary testing (95% versus 93%, p=0.24).
Comments:

Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
</table>
</td>
</tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results
					<table><tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear Unclear Risk of Bias</td></tr><tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear Unclear Risk of Bias</td></tr></table>
Lester et al (2003)					
Retospective Study Laboratory Setting	To establish the impact of the All Wales Lymphoma Panel review on clinical management decisions	N=99 cases for whom submitted diagnosis was changed as a result of central pathological review	Cases submitted for central review Hypothetical management plan created within MDT using the original submitted diagnosis and other patient information Each patient was presented and discussed as if a new referral and MDT members were not told that the cases used the original diagnoses to minimise bias	Actual management plan received by the patient	*Change in management* • 46/99 (46%) had a change in management as a result of central pathological review • 37/99 (37%) had a submitted diagnosis of a specific non-Hodgkin lymphoma entity reclassified to another NHL entity on review but of these only 6 (16%) resulted in a change in management. • 29/99 (29%) of cases resulted in a change in diagnosis from lymphoma to reactive lymphadenitis and 18/29 (62%) had a change in management as a result. • 13/99 (13%) of original reactive lymphadenitis diagnoses were reclassified as a specific lymphoma entity on review and 10/13 had a change in management as a result. • 7/99 (7%) of cases had a submitted diagnosis of Hodgkin's lymphoma reclassified to a specific NHL entity on review resulting in a change in management for 6/7 cases. • 6/99 (6%) cases with a submitted diagnosis of a specific NHL entity were reclassified to Hodgkin's lymphoma on review resulting in a change in management for 3/6 patients. • In 6/99 (6%) of cases a submitted lymphoma entity diagnosis was reclassified to another non-haematological malignancy on review and resulted in a change in management in 2 cases. • 1/99 (1%) case was reclassified from another specific non-haematological malignancy to a specific lymphoma entity and resulted in a change in management. *Treatment to No Treatment* • 43% of management changes resulted in a 'treatment to no treatment' decision • 22% of management changes resulted in a 'no treatment to treatment' decision with patients receiving oncological treatment in 9/10 cases. • 35% (n=16) patients had a 'change in oncological treatment' as a result of review, with 13/16 patients receiving a change in chemotherapy regimen. Specialist central pathological review impacted on patient management in three key areas: • Inappropriate oncological treatment • Unnecessary oncological treatment • Delay in oncological treatment *Comment* A change in management was diagnosed as: • Treatment to no treatment • No treatment to treatment • Change in oncological treatment

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Comments:</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Yes (UK study)

Low Risk of Bias</td></tr>
</table>
</td></tr>
<tr><td colspan="6">Matasar et al (2012)</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

1 January 2001 to 30 June 2001 and 1 January 2006-30 June 2006</td><td>To test the hypothesis that increased familiarity with the WHO classification of haematological malignancies is associated with a change in frequency of major diagnostic revision at pathology review.</td><td>N=719

Jan 2001-June 2001 N=365
Jan 2006-June 2006 N=354

There was a predominance of white, non-Hispanics and a younger median age when compared with population-based statistics (SEER)</td><td>Diagnosis and review in 2001 using the WHO classification of haematologic al malignancies</td><td>Diagnosis and review in 2006 using the WHO classification of haematological malignancies</td><td>• Agreement between the submitted and review diagnosis (most recent diagnosis was considered the submitted diagnosis)
• Factors associated with the rate of major diagnostic revisions

Agreement between the submitted and review diagnosis (most recent diagnosis was considered the submitted diagnsosis)
Agreement
Minor Discrepancy (would result in a different diagnosis but would not alter management according to NCCN guidelines)
Major Discrepancy (those that would alter management according to guidelines published by the NCCN)

Factors associated with the rate of major diagnostic revisions
Available patient demographic data (age, gender, race and ethnicity)
Clinical features (original diagnosis, type of biopsy, site of biopsy, immunohistochemistry reviewed or carried out at MSKCC, additional biopsy, type of referring lab)

Pathology review resulted in a major revision in 17.8% of cases in 2001 and in 16.4% of cases in 2006 (p=0.6)</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td>Hodgkin lymphoma was over represented in comparison with population based statistics
T-cell lymphomas increased from 2001 to 2006 which was temporally associated with the development of a focused T-cell lymphoma program giving an imbalance in the distribution of referring diagnoses between the two time periods (p=0.007).</td><td></td><td></td><td>
<table>
<tr><th>Diagnostic Revision</th><th>2001(n=365)</th><th>2006 (n=355)</th><th></th></tr>
<tr><th></th><th>N (%)</th><th>N (%)</th><th>P</th></tr>
<tr><td>Major Diagnostic Revision</td><td></td><td></td><td></td></tr>
<tr><td>MSKCC or other NCI-CCC secondary review</td><td>78 (21.4)</td><td>66 (18.6)</td><td>0.35</td></tr>
<tr><td>MSKCC revision of submitted diagnosis</td><td>65 (17.8)</td><td>58 (16.4)</td><td>0.60</td></tr>
<tr><td>Prior NCI-CCC revision (MSKCC confirmed)</td><td>13 (3.6)</td><td>8 (2.3)</td><td></td></tr>
<tr><td>Minor Diagnostic Revision</td><td></td><td></td><td></td></tr>
<tr><td>MSKCC or other NCI-CCC secondary review</td><td>24 (6.6)</td><td>31 (8.7)</td><td></td></tr>
<tr><td>MSKCC revision of submitted diagnosis</td><td>24 (6.6)</td><td>31 (8.7)</td><td></td></tr>
<tr><td>Prior NCI-CCC revision (MSKCC confirmed)</td><td>0 (0)</td><td>0 (0)</td><td></td></tr>
<tr><td>No Diagnostic Revision</td><td>263 (72.1)</td><td>258 (72.7)</td><td></td></tr>
</table>
<table>
<tr><th>Original Diagnosis</th><th>Revised Diagnosis</th><th>2001, number revised (% of original)</th><th>2006 number revised (% of original)</th></tr>
<tr><td>Benign</td><td>Lymphoma (any)</td><td>3/6 (50)</td><td>1/5 (20)</td></tr>
<tr><td>Lymphoma (any)</td><td>Benign</td><td>1/330 (0.3)</td><td>6/333 (2)</td></tr>
<tr><td>Non-diagnostic/ambiguous</td><td>Diagnostic/definitive</td><td>26/72 (36)</td><td>25/46 (54)</td></tr>
<tr><td>Diagnostic/definitive</td><td>Non-diagnostic/definitive</td><td>13/260 (5)</td><td>12/310 (4)</td></tr>
<tr><td>HL</td><td>NHL</td><td>3/72 (4)</td><td>2/57 (4)</td></tr>
<tr><td>NHL</td><td>HL</td><td>1/251 (0.4)</td><td>1/275 (0.3)</td></tr>
<tr><td>Classical HL</td><td>Nodular Lymphocyte Predominant Hodgkin Lymphoma</td><td>1/69 (1)</td><td>1/51 (2)</td></tr>
<tr><td>T-cell neoplasm</td><td>B-cell neoplasm</td><td>3/22 (14)</td><td>2/43 (5)</td></tr>
<tr><td>Highly aggressive B-cell neoplasm</td><td>Aggressive B-cell neoplasm</td><td>2/5 (40)</td><td>3/7 (43)</td></tr>
<tr><td>Aggressive B-cell neoplasm</td><td>Highly aggressive B-cell neoplasm</td><td>3/92 (6)</td><td>0/93 (0)</td></tr>
<tr><td>Aggressive B-cell neoplasm</td><td>Indolent B-cell neoplasm</td><td>6/92 (6)</td><td>3/93 (3)</td></tr>
</table>
</td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results

Indolent B-cell neoplasm	Aggressive B-cell neoplasm	16/118 (14)	8/118 (7)
Highly aggressive B-cell neoplasm	Highly aggressive B-cell neoplasm	0/5 (0)	1/7 (14)
Aggressive B-cell neoplasm	Aggressive B-cell neoplasm	0/92 (0)	1/93 (1)

Multivariate analysis of relationship between clinical features and major diagnostic revision

Clinical Feature	Adjusted Odds Ratio (95% CI)	Adjusted P value
Biopsy site		
Lymph node	1	0.27
Skin	1.44 (0.76-2.75)	
Other	0.73 (0.44-1.19)	
IHC carried out at MSKCC		
No	1	0.04
Yes	1.58 (1.03-2.41)	
Referring Diagnosis		
B-cell neoplasms	1	
T-cell neoplasms	1.50 (0.76-2.94)	<0.001
Non diagnostic	2.24 (1.11-4.55)	0.03
Hodgkin Lymphoma	0.37 (0.17-0.78)	0.009
Rare Diagnosis	3.52 (1.37-9.09)	0.009
Year of Pathology Review		
2001	1	
2006	0.84 (0.56-1.26)	0.4

Comment:

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>

Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Unclear Risk of Bias</td></tr>
</table>

</td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results
Norbert-Dworzak et al (2008) Europe (Germany, Italy, Austria)					
Prospective Review Laboratory Setting	To investigate whether flow cytometric assessment of minimal residual disease can be reliably standardised for multi-centric application	N=413 patients with acute lymphoblastic leukaemia (Centre 1=110, Centre 2=88, Centre 3=61, Centre 4=154) N=395 patients with blood and bone marrow samples received at diagnosis and from follow-up during induction treatment: PB at day 8, 15, 22, and 33; BM at day 15, 33 and 78). List Mode Data Exchange N=31 patients were selected for comparisons between centres with a total of 202 samples from 7 time points submitted to all centres for blinded LMD file interpretation.	Flow Cytometry according to a standardised process which included: Standardised SOPs for sample preparation and staining Standardisation of monoclonal antibodies for manufacturer, clone and partly for flurochrome Monoclonal antibodies were strategically assorted to fixed quadruple combinations of those markers which have been proven highest relevance for MRD studies in ALL Quality Control Immunophenotyping at diagnosis	Results from each centre following standard protocol	• Qualitative Concordance of Analyses of Exchanged List-Mode Data • Quantitative Concordance of Analyses of Exchanged List-Mode Data • Concordance of Risk Estimates upon Analyses of Exchanged List-Mode Data • Reproducibility in Inter-Laboratory Sample Exchange • Agreement of MRD Results from independent patient cohorts ***Qualitative Concordance of Analyses of Exchanged List-Mode Data*** • 106/202 (53%) submitted samples were classified as MRD positive and 96 as negative • Observed versus expected agreement was 89%, 97%, 93% and 96% for each of the centres • All four of the centres agreed on MRD status of samples in 76% of cases overall and in 78% of MRD positive and 73% of MRD negative samples. • There was no significant difference between sample series1 (n=15 patients recruited in early 2002) and series 2 (n=16 patients recruited in late 2003). • Agreement by at least 3 of the centres was found in 96% of the total sample cohort • Reasons for discordance included disturbance by normal lymphoid regeneration (n=3) MRD at the limits of detection (n=2) and technical flaws (n=3). • Agreement was best in bone marrow samples from day 15 (86% by four centres) and day 78 (81%). Samples from day 33 had lowest agreement (52%). 3 centres agreed in 100%, 96% and 84% of cases respectively • In analysing peripheral blood samples from days 0, 8, 15 and 33 there was complete agreement between centres in 100%, 83%, 62% and 73% respectively (by 3 centres it was at least 97% at all time points) • According to leukaemia phenotype, agreement was 78% in samples from BCP-ALL and 66% in T-ALL samples (at least 3 centres agreed in 96% and 94% respectively) ***Quantitative Concordance of Analyses of Exchanged List-Mode Data*** • Overall concordance of observed versus expected MRD-values was high (ICC=0.979) (series 1 ICC=0.986 and series 2 ICC=0.975) • There was little variance between centres 1 to 4 regarding their agreement in their observed and expected votes (Centre 1 ICC=0.983; Centre 2 ICC=0.993, Centre 3 ICC=0.997, Centre 4 ICC=0.995) • The variance in the ability to interpret data in relation to sample origin was small ((Centre 1 ICC=0.987; Centre 2 ICC=0.993, Centre 3 ICC=0.922, Centre 4 ICC=0.997) • In MRD positive samples (n=106), correct MRD levels were recorded by centres 1 to 4 in 82%, 93%, 85% and 94% respectively. All centres were in agreement in 67% of samples and at least 3 centres were in agreement in 86% of samples. • Concordance was slightly better between centres for bone marrow samples compared with blood samples with all 4 centres in agreement in 72% of bone marrow samples compared with 56% of blood samples. • Level of agreement declined with the level of MRD, samples positive ≥10% (n=27), ≥1-10% (n=21), ≥0.1-1% (n=33) and <0.1% (n=25) showed agreement in all four centres in 96%, 71%, 64% and 36% respectively. • Cumulatively there were 25 false-negative estimates (6%) among 420 available single values from all positive samples and an additional 22 estimates (5.2%) described the wrong levels of MRD. • Among the 96 negative samples, concordantly negative votes were given in 74% and by at least 3 centres in 97%. There were 24 false positive estimates (6.3%).

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td>Continued training of study group members</td><td></td><td>
Concordance of Risk Estimates upon Analyses of Exchanged List-Mode Data
• Observed risk estimates matched expected in 79%, 89%, 100% and 93% of centres respectively (based on the double time point risk algorithm) and matched expected in 96%, 89%, 100% and 89% of centres (based on the single time point algorithm (day 15, bone marrow)).

Reproducibility in Inter-Laboratory Sample Exchange
• 63 samples were exchanged between two centres with a positive/negative concordance of 90% (κ=0.81)
• The reproducibility of MRD values including quantitative aspects was high (ICC=0.97 for relative estimates)
• Concordance in the artificial dilution experiments was high between all four centres (ICC=0.98)
• Of 164 MRD values available (from 42 submitted samples) sensitivity was 95.6% and specificity was 90.2%
• MRD-status agreement was 77% (samples with poor agreement was due to insufficient red cell lysis after prolonged transportation or too few sample resulting from tube leakage)

Agreement of MRD Results from independent patient cohorts
• Agreement between the four centres with respect to available MRD results from their locally recruited patient cohorts did not differ significantly at the various time points for blood samples. In bone marrow analysis agreement between the centres differed significantly only at day 15 (p<0.001) and overall agreement was 89%.
• The proportions of patients distributed to each risk group did not differ significantly

Comments:

Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Unclear Risk of Bias</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td colspan="6">Norgaard et al (2005) Denmark (free, tax-supported health care)</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

January 1994 – December 1999</td><td>To examine the data quality and quantifying the impact of any misclassification of the diagnoses on the survival estimates</td><td>N=1159 patients identified in 2 registries (Danish Cancer Registry (DCR) and North Jutland Hospital Discharge Registry(HDR))

Inclusion
Patients registered for the first time with a haematological malignancy discharge diagnosis during 1994-1999

Exclusion
Patients <15 years
Patients who were registered prior to 1994 with an haematological diagnosis based on ICD-8</td><td>Danish Cancer Registry (DCR)</td><td>North Jutland Hospital Discharge Registry</td><td>

Degree of completeness
Positive Predictive Value
Survival

Degree of Completeness
PPV (defined as the proportion of patients registered with a haematological malignancy in HDR and in DCR
Survival

78.3% (n=908) of patients were found in both registries, 14.4% (n=167) were found in the HDR registry only and 7.3% (n=84) were found in the DCR only

Degree of Completeness and Positive Predictive Value

Completeness overall was 91.5% (95% CI 89.6%-93.1%)
PPV was 84.5% (95% CI 82.2%-86.5%) when using the DCR as reference standard

<table>
<tr><th></th><th colspan="6">Patients Registered</th></tr>
<tr><th></th><th>Both registries (%)</th><th>HDR (%)</th><th>DCR (%)</th><th>Total</th><th>Degree of Completeness</th><th>PPV (95% CI)</th></tr>
<tr><td>All haematological malignancies</td><td>908 (78.3)</td><td>167 (14.4)</td><td>84 (7.3)</td><td>1159</td><td>91.5 (89.6-93.1)</td><td>84.5 (82.2-86.5)</td></tr>
<tr><td>Acute Myeloid Leukaemia</td><td>73 (62.4)</td><td>35 (29.9)</td><td>9 (7.7)</td><td>117</td><td>89 (80.4-94.1)</td><td>67.6 (58.3-75.7)</td></tr>
<tr><td>Hodgkin’s disease</td><td>55 (65.5)</td><td>22 (26.2)</td><td>7 (8.3)</td><td>84</td><td>88.7 (78.5-94.4)</td><td>71.4 (60.5-80.3)</td></tr>
<tr><td>Non-Hodgkin’s lymphoma or chronic lymphocytic leukaemia</td><td>523 (76.6)</td><td>90 (13.2)</td><td>70 (10.3)</td><td>683</td><td>88.2 (85.3-90.6)</td><td>85.3 (82.3-87.9)</td></tr>
<tr><td>Multiple Myeloma</td><td>130 (76)</td><td>28 (16.4)</td><td>13 (7.6)</td><td>171</td><td>90.9 (85.1-94.6)</td><td>82.3 (75.6-87.4)</td></tr>
</table>
Pathological Record Reviews

73.8% of patients registered in DCR only were confirmed as having a correct or most likely correct diagnosis compared with 42.5% for patients registered in HDR only (histopathology or peripheral blood smears).
96/1075 (8.9%) of patients with a haematological malignancy registered in HDR could not be confirmed as actually having a haematological malignancy and HDR missed 62 patients who were confirmed as correctly diagnosed in DCR.
71 patients registered in HDR only, actually had a haematological malignancy
992 patients were registered in DCR as having a haematological malignancy giving an under-notification in DCR by approximately 7%.

Survival
<table>
<tr><th></th><th>Mortality Rate</th><th>95% CI</th></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>
<table>
<tr><th></th><th>Ratio (MR)</th><th></th></tr>
<tr><td>All haematological malignancies</td><td>0.98</td><td>0.88-1.09</td></tr>
<tr><td>Acute Myeloid Leukaemia</td><td>0.91</td><td>0.67-1.24</td></tr>
<tr><td>Hodgkin’s disease</td><td>1.33</td><td>0.77-2.38</td></tr>
<tr><td>Non-Hodgkin’s lymphoma or chronic lymphocytic leukaemia</td><td>0.98</td><td>0.84-1.14</td></tr>
<tr><td>Multiple Myeloma</td><td>0.87</td><td>0.68-1.12</td></tr>
</table>

In acute myeloid leukaemia and in multiple myeloma HDR overestimated the survival by 10-15% while in Hodgkin’s disease survival was underestimated by 33% compared with DCR.
Survival of patients registered in DCR only was around 20% lower than survival of patients registered in both DCR and HDR
Survival of patients registered in HDR only was around 10% lower than survival of patients registered in both DCR and HDR
Differences in survival were most pronounced in the period immediately following diagnosis

Comments

ICD-9 was never used in Denmark
Reporting to the Danish Cancer Registry became mandatory for all doctors in 1987
Patients recorded in both registries were considered to be correctly diagnosed.

</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear (free, tax-supported health care)

Unclear Risk of Bias</td></tr>
</table></td></tr>
<tr><td colspan="6">Proctor et al (2011) UK</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

2003-2008</td><td>A large scale assessment of expert central review in a UK regional cancer network and the impact of discordant diagnoses on patient management as well as the financial and educational implications of providing a centralised service.</td><td>N=1949 samples sent for expert central review

N=1873 (96.1%) were received with a primary diagnosis

Patient pathology samples sent for central expert review over a 6 year period Patient samples without a primary diagnosis were included but analysed separately</td><td>Expert Review</td><td>Initial Diagnosis</td><td>• Concordance
The overall discordance rate was 27.4% (513/1873) though the rate differed significantly between different diagnoses.

Table: Concordant and Discordant Diagnosis in the 10 most common lymphoid malignancies
<table>
<tr><th></th><th colspan="10">Expert/Final Pathology</th></tr>
<tr><th>Referral Pathology</th><th>DLBL</th><th>FL</th><th>PCN</th><th>cHL</th><th>CLL</th><th>LPL</th><th>Reactive</th><th>MCL</th><th>MZL</th><th>TCL</th></tr>
<tr><td>DLBL</td><td>361*</td><td>7</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>2</td><td>1</td><td>1</td></tr>
<tr><td>FL</td><td>10</td><td>242*</td><td>0</td><td>0</td><td>2</td><td>3</td><td>0</td><td>1</td><td>2</td><td>0</td></tr>
<tr><td>PCN</td><td>0</td><td>0</td><td>187*</td><td>0</td><td>0</td><td>3</td><td>2</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>cHL</td><td>0</td><td>1</td><td>0</td><td>172*</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>CLL</td><td>1</td><td>6</td><td>0</td><td>0</td><td>139*</td><td>5</td><td>0</td><td>3</td><td>1</td><td>0</td></tr>
<tr><td>LPL</td><td>1</td><td>4</td><td>1</td><td>0</td><td>0</td><td>53*</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>Reactive</td><td>1</td><td>4</td><td>0</td><td>1</td><td>0</td><td>2</td><td>33*</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>MCL</td><td>1</td><td>1</td><td>0</td><td>0</td><td>4</td><td>0</td><td>0</td><td>29*</td><td>0</td><td>0</td></tr>
<tr><td>MZL</td><td>2</td><td>7</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>24*</td><td>0</td></tr>
<tr><td>TCL</td><td>3</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>61*</td></tr>
<tr><td>Burkitts</td><td>3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
</table></td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results

Lymphoma										
Unspecified Lymphoma	47	42	4	4	25	14	2	7	6	6
Low-grade Lymphoma	0	0	0	0	0	1	0	0	0	0
High-grade Lymphoma	63	5	0	0	0	0	0	0	0	0
Normal/no lymphoma	0	1	1	0	1	0	2	1	0	0
Other	0	0	1	1	0	2	1	0	0	1
Total Samples	512	333	195	185	175	88	47	44	37	70
Discordant samples (%)	132 (25.8)	78 (23.4)	7 (3.6)	7 (3.8)	35 (20)	30 (34.1)	8 (17)	15 (34.1)	10 (27)	9 (12.9)
No diagnosis provided (%)	19 (3.7)	13 (3..9)	1 (0.5)	6 (3.2)	1 (0.6)	5 (5.7)	2 (4.3)	0 (0)	3 (8.1)	10 (14.3)

Discordance rates varied significantly over time with 32%discordance in 2003 dropping to between 13% and 15% after 2006.

350/512 discordant diagnoses were assessed to see whether expert panel review would have altered treatment and it was noted that expert panel review would have resulted in a significant change in 11% 9n=39) patients and in 39% (n=136) central review would have led to minimal changes to patients care.
In 50% (n=175) of patients, the primary diagnosis provided insufficient or outdated information and, without central review, would have led to delayed or potentially inappropriate treatment.

Comments
Pathologic discordance was defined as a disagreement between the primary or referred diagnosis and the diagnosis recorded after expert review

Diagnoses not conforming to the WHO system were considered discordant

Primary diagnoses were not considered discordant if they failed to provide additional details relating to grade or subtype

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Yes

Low Risk of Bias</td></tr>
</table></td></tr>
<tr><td colspan="6">Rane et al (2014) India</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

March 2011 – no end date reported</td><td>To evaluate the ability and interobserver variability of pathologists with varying levels of experience and with an interest in lymphomas to diagnose Burkitt Lymphoma in a resource limited set up.</td><td>N=25 cases selected

Diagnosis of Burkitt Lymphoma based either on clinical features, morphological features and immunophenotypes</td><td>Consensus Diagnosis</td><td>Initial Independent Assessment</td><td>• Initial Independent Assessment
• Interobserver variation in morphological features
• Parameters used to differentiate between classic CL, atypical BL and B-cell lymphoma intermediate between Burkitt’s and DLBL
• Consensus Diagnosis
• Concordance with consensus diagnosis
• Effect of tissue fixation, age group and provision of additional information on revision of diagnoses
• Accuracy of pathologists
• Sensitivity and Specificity to diagnose Burkitt Lymphoma

Initial Independent Assessment
• 10 pathologist committed to a diagnoses in all 25 cases while 3 pathologists committed to a diagnosis in 24/25 cases, 1 pathologist committed in 23/25 cases.
• There was poor concordance for independent diagnosis (κ=0.168, SE±0.018)
• Level of experience showed direct correlation with expert lymphoma pathologists showing marginally higher concordance rates (κ=0.373, SE±0.071) and general pathologists showing the lowest (κ=0.138, SE±0.035)

Interobserver variation in morphological features
• There was very low concordance for morphological features tested among all pathologists (κ=0.192, SE±0.05) and concordance for morphological diagnosis was highest among expert lymphoma pathologists (κ=0.356, SE±0.127).
• Highest concordance rate was observed for nuclear contour (κ=0.896, SE±0.110) and was lowest for nuclear</td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results
					prominence (κ=-0.62, SE±0.124) *Parameters used to differentiate between classic CL, atypical BL and B-cell lymphoma intermediate between Burkitt's and DLBL* • Cross tabulation of morphological and immunohistochemical features against the independent final diagnosis showed that pathologists were least likely to accept deviation from certain features perceived to be characteristics of Burkitt Lymphoma (intermediate cell size, CD10 + MIB-1 labelling of greater than 90% and the greater the deviation the more likely a pathologist was to classify the case as either atypical BL or B cell lymphoma intermediate between Burkitt's and DLBL. *Consensus Diagnosis* • 12/14 pathologists attended the consensus meeting and a consensus was reached in 23/25 cases, unanimously in 19 cases and consensus based (≥8 pathologists in agreement) in 4 cases. • Level of agreement between pathologists for revised diagnosis was very high (κ=0.835, SE±0.021) and was similar across the different groups of pathologists • Revision of diagnosis was highest amongst general pathologists and lowest among lymphoma experts (p=0.121) • Revision was highest for cases originally diagnosed as either atypical BL or B cell lymphoma intermediate between Burkitt's and DLBL. and minimum revision occurred in classic BL (p=0.001). *Concordance with consensus diagnosis* • Concordance of independent diagnosis and consensus diagnosis was low and highly variable (κ=0.259, SE±0.039; median 0.207; range -0.131-0.667). • Concordance with independent diagnosis increased and variability decreased with increasing experience of diagnosing lymphomas • Concordance of the revised diagnosis with consensus diagnosis was high (κ=0.633, SE±0.011, median 0.656) *Effect of tissue fixation, age group and provision of additional information on revision of diagnoses* • No difference was observed in the distribution of fixation and staining scores across the diagnostic categories (p=0.654) • Equal proportions of cases were reclassified in all three grades of fixation: (means Grade 1=54.167±29.167,Grade 2= 47.222±7.217 and Grade 3=50±6.989; p=0.931).C-MYC status, EBER-ISH results and BCL6 IHC results did not affect the frequency of revision of diagnoses • Age of patients (adult versus paediatric) did not affect the rates of revision of diagnosis (mean revision 45.513±6.579% in patients <18 years and 53.472±7.429 in adult patients. *Accuracy of pathologists* • Expert lymphoma pathologists were significantly more likely to make a correct diagnosis compared with both the pathologists with experience (OR=3.14, p=0.012) and the general pathologists (OR=5.3, p=0.00032) and pathologists with experience were more likely to make a correct diagnosis compared with general pathologists though this was not statistically significant (OR=1.69, p=0.062). • Mean change of accuracy by IHC over morphology was 9.698±4.799 and mean change of accuracy by discussion/consensus meeting over that by IHC was 47.464±5.039%.

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>
<table>
<tr><th>Mean Accuracy</th><th>Morphological diagnosis</th><th>Morphological Diagnosis + IHC</th><th>Revised Diagnosis post consensus meeting</th><th>Burkitt Lymphoma group</th><th>DLBL</th><th>Atypical BL</th><th>B-cell lymphoma intermediate between BL and DLBL</th></tr>
<tr><td>All</td><td>36.79±2.631 %</td><td>45.963±13.825 %</td><td>95.652±1.31 1%</td><td>72.619±7.5 36%</td><td>58.9 28±8 .535 %</td><td>24.186±7 .026%</td><td>35.714±10.16 6%</td></tr>
<tr><td>Expert lymphoma pathologists</td><td>~42%</td><td>66.667±13.825 %</td><td>97.101±2.89 8%</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Pathologists with lymphoma experience</td><td></td><td>51.087±4.82%</td><td>92.391±2.73 5%</td><td></td><td></td><td></td><td></td></tr>
<tr><td>General Pathologists</td><td>~33%</td><td>34.161±3.727 %</td><td>97.391±1.46 9%</td><td></td><td></td><td></td><td></td></tr>
</table>
Sensitivity and Specificity to diagnose Burkitt Lymphoma

Expert lymphoma pathologists had the highest sensitivity (96.8%) and specificity (94.44%) for the diagnosis of Burkitt Lymphoma (typical and atypical)
General pathologists had a higher sensitivity (78.57% versus 65.63%) compared with pathologists with lymphoma experience, however pathologists with lymphoma experience had much higher specificity (80.56% versus 63.49%).

Comments
The 14 pathologists were divided into three groups:

A1-A3 expert lymphoma pathologists working in diagnostic centres with >500 lymphoma cases/year,
B1-B4 pathologists with experience in lymphomas working in general hospitals with some training in lymphoma
C1-C7 Other pathologists involved in diagnostic surgical pathology

</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Low Risk of Bias</td></tr>
</table></td></tr>
<tr><td colspan="6">Siebert et al (2001) USA</td></tr>
<tr><td>Retropsective Study

Laboratory Setting

July 1995- December 1997</td><td>To compare diagnoses made at a community and an academic centre to evaluate the reproducibility of the revised European-American Classification</td><td>N=188 lymphoid neoplasms subtyped according to revised European-American classification criteria</td><td>Review of community hospital assessments at an academic centre</td><td>lymphoid neoplasms subtyped according to revised European-American classification criteria at a community hospital</td><td>• Concordance
Subtype was concordant for 88.8% of cases (167/188)

Methods used for diagnosing and subtyping
<table>
<tr><th>Method</th><th>Frequency</th><th>Concordant</th><th>Discordant</th></tr>
<tr><td>Morphologic Examination</td><td>7 (3.7)</td><td>7 (3.7)</td><td>0 (0)</td></tr>
<tr><td>Morphologic Examination and paraffin-section immunohistochemical examinations</td><td>49 (26.1)</td><td>41 (21.8)</td><td>8 (4.3)</td></tr>
<tr><td>Morphologic Examination and paraffin-section immunohistochemical examinations and flow cytometry</td><td>57 (30.3)</td><td>48 (25.5)</td><td>9 (4.8)</td></tr>
<tr><td>Morphologic Examination and flow cytometry</td><td>75 (39.9)</td><td>71 (37.8)</td><td>4 (2.1)</td></tr>
<tr><td>Total</td><td>188 (100)</td><td>167 (88.8)</td><td>21 (11.2)</td></tr>
</table>
Additional Data/material provided for academic centre review before diagnosis of 44 cases
<table>
<tr><th>Method</th><th>Frequency</th><th>Concordant</th><th>Discordant</th></tr>
<tr><td>Additional Clinical or Laboratory Data</td><td>10</td><td>7</td><td>3</td></tr>
<tr><td>Paraffin embedded tissue</td><td>18</td><td>13</td><td>5</td></tr>
<tr><td>Flow cytometry histograms</td><td>22</td><td>19</td><td>3</td></tr>
<tr><td>Cytogenetic or molecular test results</td><td>2</td><td>1</td><td>1</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>
Comments
For each case, clinical data, glass slides for morphologic evaluation and immunophenotying data were submitted for blinded review at an academic centre.

Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td>No – do not compare services in terms of whether they are co-located or networked.</td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Low Risk of Bias</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td colspan="6">Stevens et al (2012)</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

January 2006 – May 2010</td><td>To observe concordance and discrepancies between local findings and the specialist opinion.</td><td>N=125 patients visiting the Hodgkin outpatient clinic

Newly diagnosed and previously untreated patients with HL</td><td>Central Review</td><td>Regional/Community Hospital Review</td><td>
Pathology
Staging
Therapy
Pathology
There was agreement in 108/125 (86%) of cases between the pathologists in the referring hospital and the RUN MC; minor discordances were recorded in 12 cases and major discordance was recorded in 5 cases.

<table>
<tr><th></th><th colspan="5">Referring hospital</th></tr>
<tr><th>Central Review</th><th>NScHL</th><th>MCcHL</th><th>LRcHL</th><th>NLPHL</th><th>NOS</th></tr>
<tr><td>NScHL</td><td>75</td><td>3</td><td></td><td></td><td>4</td></tr>
<tr><td>MCcHL</td><td></td><td>10</td><td></td><td></td><td>1</td></tr>
<tr><td>LRcHL</td><td></td><td></td><td>5</td><td>1</td><td></td></tr>
<tr><td>NLPHL</td><td></td><td></td><td>2</td><td>10</td><td></td></tr>
<tr><td>NOS</td><td>1</td><td>1</td><td></td><td></td><td>8</td></tr>
<tr><td>Others</td><td></td><td></td><td>1</td><td>1</td><td></td></tr>
</table>
Staging
The Ann Arbor stage could be attributed to 123/125 cases (98%) of patients at central review and 95/123 (77%) were concordant with regional results.
There were 10 minor discordant and 18 major discordant results; discordant results included downscaling or upscaling after central review.

<table>
<tr><th></th><th colspan="8">Ann Arbor Referring Hospital</th></tr>
<tr><th>Ann Arbor Centralised Revision</th><th>Stage I (favourable)</th><th>Stage I (unfavourable)</th><th>Stage II (favourable)</th><th>Stage II (unfavourable)</th><th>Stage III (good risk)</th><th>Stage III (poor risk)</th><th>Stage IV (good risk)</th><th>Stage IV (poor risk)</th></tr>
<tr><td>Stage I (favourable)</td><td>9</td><td></td><td>2</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Stage I (unfavourable)</td><td></td><td>4</td><td></td><td>2</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Stage II (favourable)</td><td>4</td><td></td><td>21</td><td>1</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>Stage II (unfavourable)</td><td></td><td>1</td><td>6</td><td>26</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Stage III (good risk)</td><td></td><td></td><td>1</td><td>1</td><td>14</td><td></td><td></td><td></td></tr>
<tr><td>Stage III</td><td></td><td></td><td></td><td>4</td><td></td><td>6</td><td></td><td></td></tr>
</table>
</td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results

(poor risk)								
Stage IV (good risk)	1		1	1			5	
Stage IV (poor risk)				1		1		10
Missing/Other					1		1	

Therapy
Central treatment advice could be given in 124/125 cases
Regional centres had already defined treatment in 104 cases and treatment advice was concordant in 84/104 (81%) of cases
Central review led to treatment changes in 20/104 (19%) of cases based on either changes in pathology or staging results. Treatment changes included changes to radiotherapy field, changes in the number of ABVD cycles and involved nodes for the IN-RT and other changes including changes in chemotherapy.

	Referring Hospital							
Central Revision	**IF-RT**	**ABVDx 6**	**ABVDx 8**	**ChIVPPx6 -8**	**Other Chemo**	**ABVDx3 + IN-RT**	**ABVD x4 + IN-RT**	**Missing Data**
IF-RT	8					1		
ABVDx6		27				2	3	6
ABVDx8			1					
ChIVPPx6-8				1				1
Other Chemo	1				2			2
ABVDx3 + IN-RT	2	1				22	1	2
ABVD x4 + IN-RT						5	23	7
Missing Data		1						
Other Chemo+RT						1	1	1
Other Treatment							2	1

Comments

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td></td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Low Risk of Bias</td></tr>
</table></td></tr>
<tr><td colspan="6">Strobbe et al (2014) The Netherlands</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

2000-2001
2005-2006</td><td>To investigate whether implementation of an expert panel led to better quality of initial diagnoses by comparing the rate of discordant diagnoses after the panel was established compared with discordance rate 5 years later

To evaluate whether lymphoma types with high discordance rate could be identified</td><td>N=161 referred to the expert panel
N=183 reviewed at a later date

2000-2001

N=433 patients with a diagnosis of malignant lymphoma

N=89 patients excluded (not possible to retrieve pathology, tissue, diagnosis at autopsy, fine needle aspiration only, patients already sent for consultation, cutaneous lymphoma)</td><td>Expert Panel review</td><td>Initial Diagnosis</td><td>• Discordance rate in 2000-2001
• Discordance rate in 2005-2006

• Overall discordance rate decreased from 14% in 2000-2001 to 9% in 2005-2006 (p=0.06)
• In 2000-2001, the highest rate of discordance was observed for lymphoma with transformation (90%), lymphoma NOS (61%), low grade lymphoma NOS (44%) and follicular lymphoma grade 3 (33%)
• In 2005-2006, the highest rate of discordance was observed for Lymphoma NOS (57%), lymphomas with transformation (56%), follicular lymphoma grade 3 (50%) and nodular lymphocyte predominant Hodgkin lymphoma (50%)
• Despite overall decrease in discordance, 3/4 groups with the highest discordance rates were the same
In 2000-2001, 11% of cases were discordant compared with 16% who were not referred (p=0.2) and in 2005-2006, discordance rate for referred versus non-referred were 10% versus 9% (p=0.8).

Comments
• All seven hospitals in the region agreed to submit histological slides of all new cases of patients with a diagnosis of malignant lymphoma
• Initial diagnosis was made in three pathology laboratories
• Expert panel consisted of three expert haematopathologists (one from each laboratory) so haematopathologists sometimes reviewed their own cases (no information as to whether this was blinded review though reviewers were not blinded to initial diagnosis) but the other two reviewers confirmed/rejected the diagnosis.</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td>N=344 cases included in the analysis

2005-2006

N=473 cases of malignant lymphoma

N=103 cases excluded (not possible to receive pathology tissue, fine needle aspiration only, diagnosed at autopsy, already sent for consultation, cutaneous lymphoma)

N= 370 cases included in the analysis</td><td></td><td></td><td>Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td></td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Low Risk of Bias</td></tr>
</table>
</td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results
Van Blerk et al (2003)					
Retrospective Study Laboratory Setting January 2000 – November 2001	To report first experiences from Belgian national external quality assessment scheme (EQAS)	N=17 blood samples were sent for evaluation by EQAS N=41 laboratories 61.5% non-university hospitals 25.6% university hospitals 12.9% private laboratories 78.4 % Sample analysis was performed within 24 hours and 96.2% within 48 hours	External Quality assessment Review (an expert laboratory tested both the fresh samples immediately after apherisis and the mailed samples)	N/A	• Stability • Intralaboratory reproducibility • Homogeneity • Interlaboratory reproducibility • Single vs. Dual Platform • Influence of Gating strategy • CD4+,CD3+ and CD8+CD3+ cells versus total CD4 and CD8 cells • Abnormal Samples *Stability* No significant difference in variation was observed over the test period Variability increased with age of sample but stability of control samples appeared satisfactory until day 2. Results between fresh and mailed samples did not differ significantly Results obtained by participants within 24 hours of blood collection and those obtained from specimens processed later *Intralaboratory Reproducibility* Within laboratory variability and relative contribution to total variability was assessed by sending duplicate samples to labs and asking them to analyse them twice. For duplicate measurements, differences ranged between -5.0 and 5.0% for the percentages of lymphocyte subsets and between -0.33 and 0.28 10^9/litre for the absolute counts. Between duplicate measurements or duplicate samples, no significant difference was observed *Homogeneity* The homogeneity of the specimens was demonstrated by the ratios of duplicate samples being practically equal to 1 *Interlaboratory Reproducibility* Between-laboratory CV values for the white blood cell and lymphocyte count ranged between 2.9-5.6% and 3.9-16.2% respectively Overall between laboraotory variability for the percentage of CD3+, CD4+, CD8+ and CD19+ cells was 4.0, 5.0, 13.2 and 16.2% respectively. Median CVs of the absolute values were 12.2, 11.4.16.4 and 16.5% for CD3+, CD4+, CD8+ and CD19+ cells respectively *Single versus dual platform approach* Overall interlaboratory CVs obtained from 2 surveys with single platform approach were 6.6% (range, 3.5-8.8%), 7.4% (range 1.6%-11.8%), 9.1% (range, 2.5-15.3%) and 17% (range, 5.6-34.3%) for the absolute CD3+, CD4+, CD8+ and CD19+ cell counts respectively (6 laboratories) Overall interlaboratory CVs obtained with dual platform approach were 9.3% (range 4.5-11.7%), 10.5% (range 8.3-13%), 11% (range 7.9-13.8% and 15.1% (range 10.5-21.1%) for the absolute CD3+, CD4+, CD8+ and CD19+ cell counts respectively (35 laboratories) No significant difference was observed between the two groups *Influence of gating strategy* There was no significant difference in different gating strategies observed

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results
					CD4+CD3+ and CD8+CD3+ cells versus total CD4 and CD8 cells The percentage of double-positive CD4+CD3+ cells and CD8+CD3+ cells was significantly lower than the percentage total CD4+ and CD8+ cells for a number of samples. The overall CVs for the percentages of CD4+CD3+ cells and CD8+CD3+ cells for the six surveys were, respectively 4.3 and 7.1% Overall CVs for the absolute numbers of CD4+CD3+ cells and CD8+CD3+ cells were 10.1% and 11.6% respectively Between laboratory variability for the determination of CD4+CD3+ cells and CD8+CD3+ cells was lower than for the measurement of total CD4+ and CD8+ cells The percentage of laboratories which reported measuring total CD4+ and CD8+ cells was 29.3% in January 2000 and dropped to 19.5% by November 2001. *Abnormal Sample* One survey included a specimen with an abnormal proportion of lymphocyte subsets Median values obtained by participating laboratories matched well with the results of the expert laboratory. Between laboratory variability for CD3, CD4 and CD8 was considerable *Comments* Two or three fresh anticoagulated whole blood sample were sent out to laboratories a total of six times for analysis. In two send outs, within laboratory variability and abnormal samples analysis were assessed: • Survey 2: To assess variability within each laboratory (duplicate samples, analysed twice) • Survey 4: To evaluate variability inherent to abnormal samples (samples sent included a sample from a patient suffering from chronic B-lymphocytic leukaemia) Laboratories were required to report • Date of receipt of sample • Date of sample analysis • Type of flow cytometer • Sample preparation technique • Source of antibodies • Gating strategy • Data analysis software
					Comments:

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td></td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Low Risk of Bias</td></tr>
</table></td></tr>
<tr><td colspan="6">Van de Schans et al (2013) The Netherlands</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

January 200 – December 2001</td><td>To evaluate the value of an expert pathology panel and report discordance rates between the diagnosis of initial pathologists and the expert panel and the effect on survival</td><td>N=391 patients diagnosed with primary malignant lymphoma

N=344 patients included

Inclusion
Patients with malignant lymphoma identified through the regional population based cancer registry

Three pathology labs including one academic performed</td><td>Expert review of diagnosis</td><td>Initial Diagnosis</td><td>• Discordance Rate

Rate of discordance

• 47% of all cases were actively referred for expert review with diffuse large B cell lymphoma the most common type to be referred (32%)

• Discordance rate was 14%; κ=0.84, 95% CI, 0.78-0.89)
• Discordance rate differed for patients referred (11%) compared with patients not referred (16%) though this was not statistically significant.
• Discordance rates varied between 11 and 23% for individual laboratories
• Patients with a discordant diagnosis were older (median age was 68 years versus 63 years) and the distribution of NHL subtypes was different; less DLBCL (9 vs. 36%), more LL NOS (9 vs 2%), more FL grade 3 (11 versus 3%), less TCL (0 versus 7%), less HL (4 versus 12%) and more L NOS (23 versus 2%).
• There was no statistically significant difference in 5 year survival between patients with a concordant diagnosis versus a discordant diagnosis (48% [95% CI 42-53%] versus 53% [95% CI 39-67%].</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td>the original diagnosis</td><td></td><td></td><td>
Comments:
55% of diagnoses were made in one laboratory which served 3 hospitals

NHL – Non Hodgkin Lymphoma
DLBCL – Diffuse large B cell lymphoma
LL NOS – low grade lymphoma not otherwise specified
FL – Follicular Lymphoma
L NOS – Lymphoma not otherwise specified
TCL – T cell lymphoma

Quality Assessment

<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td></td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Low Risk of Bias</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td colspan="6">Zhang et al (2007)</td></tr>
<tr><td>Retrospective Study

Laboratory Setting

2004-2005</td><td>To compare similarities and differences in results from participating laboratories and to identify variables which could potentially affect test results to discern variables important in test standardisation</td><td>N=38 laboratories

N=38 laboratories participated in the sample exchange and provided results

N=29 labs had results starting from a 10^{-5} dilution
N=40 labs had results starting from a 10^{-4} dilution
N=43 labs had results starting from a 10^{-3} dilution
N=43 labs had results starting from a 10^{-2} dilution
N=42 labs had results starting from a 10^{-1} dilution</td><td>Quantitative testing for BCR-ABL1</td><td>Results from different participating laboratories</td><td>• Test accuracy at different dilutions
Test accuracy at different dilutions (based on log reductions)

<table>
<tr><th></th><th>10^{-5} dilution</th><th>10^{-4} dilution</th><th>10^{-3} dilution</th><th>10^{-2} dilution</th><th>10^{-1} dilution</th></tr>
<tr><td colspan="6">All internal controls</td></tr>
<tr><td>Mean</td><td>4.45</td><td>3.52</td><td>2.58</td><td>1.536</td><td>0.667</td></tr>
<tr><td>SD</td><td>0.609</td><td>0.578</td><td>0.574</td><td>0.584</td><td>0.394</td></tr>
<tr><td>Median</td><td>4.52</td><td>3.56</td><td>2.63</td><td>1.6</td><td>0.605</td></tr>
<tr><td>Minimum</td><td>3.26</td><td>2.18</td><td>1.03</td><td>0.26</td><td>0.14</td></tr>
<tr><td>Maximum</td><td>6.30</td><td>4.71</td><td>3.7</td><td>3.0</td><td>1.70</td></tr>
<tr><td>Range</td><td>3.04</td><td>2.53</td><td>2.67</td><td>2.74</td><td>1.56</td></tr>
<tr><td>N</td><td>29</td><td>40</td><td>43</td><td>43</td><td>42</td></tr>
<tr><td>ABL1 as control</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Mean</td><td>4.149</td><td>3.06</td><td>2.09</td><td>1.1225</td><td>0.3773</td></tr>
<tr><td>SD</td><td>0.486</td><td>0.385</td><td>0.54</td><td>0.446</td><td>0.3404</td></tr>
<tr><td>Median</td><td>4.1</td><td>3.08</td><td>2.145</td><td>1.01</td><td>0.300</td></tr>
<tr><td>Minimum</td><td>3.26</td><td>2.34</td><td>1.03</td><td>0.5</td><td>0.14</td></tr>
<tr><td>Maximum</td><td>4.8</td><td>3.85</td><td>3.2</td><td>2.2</td><td>1.50</td></tr>
<tr><td>Range</td><td>1.54</td><td>1.51</td><td>2.17</td><td>1.7</td><td>1.36</td></tr>
<tr><td>N</td><td>10</td><td>14</td><td>16</td><td>16</td><td>15</td></tr>
<tr><td colspan="6">GAPDH, BCR, G6PD and B2M as control</td></tr>
<tr><td>Mean</td><td>4.61</td><td>3.77</td><td>2.875</td><td>1.782</td><td>0.8285</td></tr>
<tr><td>SD</td><td>0.616</td><td>0.401</td><td>0.351</td><td>0.427</td><td>0.3279</td></tr>
<tr><td>Median</td><td>4.58</td><td>3.78</td><td>2.8</td><td>1.755</td><td>0.71</td></tr>
<tr><td>Minimum</td><td>3.52</td><td>2.18</td><td>2.3</td><td>0.26</td><td>0.38</td></tr>
<tr><td>Maximum</td><td>6.3</td><td>4.7</td><td>3.7</td><td>3.00</td><td>1.70</td></tr>
<tr><td>Range</td><td>2.78</td><td>2.53</td><td>1.4</td><td>2.74</td><td>1.32</td></tr>
<tr><td>N</td><td>19</td><td>26</td><td>27</td><td>27</td><td>27</td></tr>
</table>

Effect of different variables on reported log reductions at different dilution (p values)
<table>
<tr><th>Dilution</th><th>Extraction Method</th><th>RT Primer</th><th>RT Enzyme</th><th>PCR Kit</th><th>Instrument</th><th>Standard Curve</th><th>Internal Control</th></tr>
<tr><td>10^{-5}</td><td>0.89</td><td>0.41</td><td>0.9</td><td>0.36</td><td>0.66</td><td>0.16</td><td>0.16</td></tr>
<tr><td>10^{-4}</td><td>0.84</td><td>0.52</td><td>0.4</td><td>0.21</td><td>0.75</td><td>0.11</td><td>0.001</td></tr>
<tr><td>10^{-3}</td><td>0.78</td><td>0.6</td><td>0.005</td><td>0.09</td><td>0.61</td><td>0.01</td><td><0.001</td></tr>
<tr><td>10^{-2}</td><td>0.39</td><td>0.42</td><td>0.08</td><td>0.07</td><td>0.48</td><td>0.05</td><td>0.001</td></tr>
<tr><td>10^{-1}</td><td>0.16</td><td>0.32</td><td>0.75</td><td>0.17</td><td>0.02</td><td>0.06</td><td><0.001</td></tr>
</table>

All Internal Controls
Mean and median results were all within 0.5 log of the known dilution (expected value) apart from 10^{-5} where it was 0.55 log</td></tr>
</table>

Study Type/Setting	Aim	Population	Intervention	Comaprison	Outcomes and results
					Standard Deviation was 0.6 log at all dilutions except from 10^{-1} where it was 0.4 log ABL1 Mean and median were ~1 log less than the known dilution value apart from 10^{-1} which was within 0.6 log of the expected value RNA Quality and cDNA Synthesis (spectrophotemtry and/or gel electrophoresis) Low yields did not appear to impact results Storage time did not impact sensitivity or accuracy of results (storage times ranged from 1-25 days) cDNA synthesis was done by reverse transcription and type of primers and enzymes used did not affect the sensitivity or accuracy Reagents for Quantitative PCR (Applied Biosystems kit and instruments, Roche quantification kit and light cycler, Ipsogen Fusion Quant kit or homebrew buffers) Different PCR kits and reagents used by the different laboratories did not impact the reported log reduction results Platforms (ABI Prism 7000, ABI Prism 7700, ABI Prism 7900, Roche LightCycler, Bio-Rad icycler) 91% of laboratories were able to amplify transcripts from samples diluted 10^{-4} and 66% were able to amplify transcripts from samples diluted at 10^{-5} irrespective of the platform or reagents used Calculation and use of the standard curve It appears the there it makes no overall difference whether laboratories use diluted RNA, cDNA, plasmid DNA or cell lines for generation of standard curves Internal Controls A number of internal controls including GUSB, ABL1, GAPDH, BCR, G6PD and B2M were used by the different laboratories (G6PD and ABL1 were the most frequent) Laboratories using BCR as their internal control appear to achieve the most accurate and sensitive results Laboratories using ABL1 showed log reduction values that were significantly different from those that used other internal controls in 4/5 dilutions tested.
					Comments: Reproducible results were defined as those that were different by less than 0.5 log in duplicate samples at dilutions as high as 10^{-4} and 10^{-5} and for duplicate samples at lower dilutions, values should be nearly identical. A 3-log reduction in BCR-ABL1 transcripts are consistent with major molecular response and a low incidence of disease progression whereas rising levels of BCR-ABL1 transcripts indicate a loss of response to treatment and may indicate relapse.

<table>
<tr><th>Study Type/Setting</th><th>Aim</th><th>Population</th><th>Intervention</th><th>Comaprison</th><th>Outcomes and results</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Quality Assessment
<table>
<tr><th>Question</th><th>Risk of bias (high, low, unclear, NA)</th></tr>
<tr><td>Patient selection – how were patients chosen for the study (e.g. consecutive or random sample)?</td><td>Not reported –

High risk of bias</td></tr>
<tr><td>Are the patients in the study representative of the PICO population</td><td>Yes (haematology patients)

Unclear Risk of Bias</td></tr>
<tr><td>Diagnostic service models – are they comparable to what is in the PICO?</td><td></td></tr>
<tr><td>Reference standard tests – did all patients receive the same tests to get the definitive diagnosis.</td><td>Unclear

Unclear risk of Bias</td></tr>
<tr><td>Blinding – are expert pathologists blinded to the initial diagnosis information</td><td>Unclear

Unclear Risk of Bias</td></tr>
<tr><td>Health care setting – is it applicable to the UK?</td><td>Unclear

Low Risk of Bias</td></tr>
</table></td></tr>
</table>