The diagnostic performance of current tumour markers in surveillance for recurrent testicular cancer: A diagnostic test accuracy systematic review

Brian D. Nicholson[a],[*], Nicholas R. Jones[a], Andrew Protheroe[b], Johnson Joseph[c], Nia W. Roberts[d], Ann Van den Bruel[a],[1], Thomas R. Fanshawe[a]

[a] Nuffield Department of Primary Care Health Sciences, University of Oxford, UK
[b] Department of Oncology, University of Oxford, UK
[c] Oxford University Hospitals Trust, Oxford, UK
[d] Bodleian Health Care Libraries, Oxford, UK

ABSTRACT

In this diagnostic test accuracy systematic review we summarise the evidence on the diagnostic accuracy of blood α-fetoprotein (AFP), human chorionic gonadotropin (HCG) and lactate dehydrogenase (LDH) in surveillance for testicular cancer recurrence in adults. We searched four electronic databases for studies that reported the diagnostic accuracy of HCG, AFP, and/or LDH in sufficient detail for sensitivity and specificity to be calculated by extracting a 2 × 2 table comparing biomarker positivity with testicular cancer recurrence. Screening, data extraction and QUADAS-2 quality assessment were completed by two independent reviewers. From 2406 studies, nine met our inclusion criteria. Eight reported data at the per-patient level. Sample sizes were small (range 5 to 449 patients) and clinical heterogeneity precluded meta-analysis. In most studies the specificity for recurrence with AFP and HCG was high (90–100%) but sensitivity was often relatively low, suggesting that many recurrences would not be detected by tumour markers alone. The diagnostic performance of LDH appears poorer. Studies were methodologically weak, with probable selection, incorporation and partial verification bias, and many studies were excluded for not reporting on recurrence-free patients. Limitations including small sample sizes, high heterogeneity, and inconsistent and incomplete reporting mean these results must be interpreted with caution. Despite inclusion of biomarkers in international surveillance guidance, there remains a lack of high quality evidence about their accuracy, optimal thresholds, and the most effective surveillance strategy in relation to contemporary investigative modalities. Higher quality research using data from modern-day follow-up cohorts is necessary to identify opportunities to reduce unnecessary testing.

1. Introduction

The incidence of testicular cancer is increasing, with age-standardized rates of 5.6 cases/100,000 men diagnosed in the United States and 7.0 cases/100,000 men diagnosed in the United Kingdom annually [1,2]. Approximately sixty percent are seminomas (peak incidence 35 years), and the rest predominantly non-seminomatous germ-cell tumours (NSGCT, peak incidence 25 years). Most patients present with tumour located within the testis (stage 1 disease), and long-term survival is high (US five-year survival > 95%, UK ten-year survival > 98%) [3,4]. Patients presenting with recurrent or metastatic disease at diagnosis can be defined by prognostic category according to the site of the primary tumour and metastases, and the tumour marker level. It remains a highly curable cancer: good and poor prognostic disease have five-year survival of 92% (NSGCTs) / 86% (seminoma) and 48% respectively [5]. The high rate of relatively young cancer survivors necessitates a review of the most efficient means of surveillance for recurrent disease.

During active surveillance following curative treatment for stage I disease, 10–20% of seminomas and 15–50% of NSGCT recur, most within two years [6]. No consensus exists on the optimal follow-up schedule to detect recurrence, but European guidelines recommend intensive multimodal follow-up including physical examination, tumour markers, chest radiograph and either Abdominal-Pelvic

Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) scans [7]. The biomarkers human choriogonadotropin (HCG), α-fetoprotein (AFP) and lactate dehydrogenase (LDH) are measured at a frequency determined by tumour stage and treatment received: approximately 3–4 times/year during the first three years and annually thereafter.

Previous reviews have discussed the diagnostic performance of these biomarkers for detecting recurrence of testicular germ cell tumours. These include a 2007 review focused on germ cell tumour surveillance outcomes including survival and quality of life [8] and more recent reviews that looked at biochemical markers for the diagnosis of germ cell tumours more generally (primary as well as recurrent disease) [6,9,10]. A striking feature of these reviews is that, although they may represent contemporary clinical opinion, only the first of these is systematic in nature, and none presents complete diagnostic accuracy data on the performance of biomarkers used to detect testicular cancer recurrence. Therefore, although biomarkers are recommended in international guidelines [7], the evidence base underpinning their use in detecting recurrence is unclear. A systematic diagnostic test accuracy review [11], rather than a potentially selective narrative overview, is lacking.

This diagnostic test accuracy review aims to clarify this issue by quantifying the accuracy of blood AFP, HCG and LDH as triage tests for further investigation for testicular cancer recurrence in adults. If sufficiently accurate, these biomarkers (alone, or in combination) could provide a cost-effective means of reducing unnecessary follow-up investigations and hospital episodes, and the radiation burden [12,13], particularly as the value of clinical examination and chest radiographs are questioned as tools to detect recurrence [14]. We also aimed to summarise the available evidence regarding cut-offs used to define a positive biomarker test for the purpose of diagnosis of testicular cancer recurrence.

2. Material and methods

The review protocol was registered in advance on the PROSPERO website, registration number CRD42017074683 [15].

We searched Embase, Medline, the Cochrane Central Register of Controlled Trials and the Science Citation Index (Web of Science) using the search terms shown in Appendix 1. The search was not restricted by study design and no language restriction was applied. However, conference abstracts were excluded, after checking that no published full article was available, on the basis that such reports were unlikely to provide sufficiently informative data for extraction for the purposes of this review. Eligible studies reported numerical information for any of the three tumour markers falling within the scope of this review (AFP, HCG and LDH) in adults during follow-up after surgery for primary testicular cancer, in relation to recurrent testicular cancer, as confirmed by physical examination, imaging or histopathology.

Studies were only included if they reported enough information to construct the full 2 × 2 table of biomarker elevation (yes/no) versus cancer recurrence (yes/no). Studies that reported data only for recurrent cases were therefore excluded. Combinations of AFP, HCG and LDH (for example, studies that had deemed either raised AFP or HCG as indicative of marker elevation) were considered. Studies that reported biomarker data only at initial diagnosis or time of surgery, but not subsequent surveillance, and prognostic studies, were also excluded. No restriction was made on the period of follow-up.

Elevated biomarkers at the time of initial diagnosis of the primary tumour are of prognostic value and form part of the International Germ Cell Consensus Classification staging criteria [5]. Biomarker levels are expected to normalise following orchiectomy and persistently elevated levels suggest there may be residual disease. For clarity, we have separated the prognostic role of biomarkers prior to treatment from their application in the diagnosis of recurrence. In this review we included only patients treated with curative intent and regarded biomarker

elevation during the surveillance period as a potential indicator of disease recurrence.

We undertook independent duplicate screening of all citations meeting the search criteria. Two reviewers (from BDN, NRJ and TRF) screened the studies initially by title and abstract, and then by full-text if required. Any disagreements were resolved by the third assessor. For the studies that met all inclusion criteria, two reviewers independently extracted the following information: age, tumour type, definition of recurrence used, information about treatment received and the patient pathway, the index test and the reference standard for confirming cancer recurrence, including the thresholds used to define marker positivity, and the performance of biomarkers as indicators of recurrence in the form of the full 2 × 2 table relating marker positivity to recurrence. Again, discrepancies were resolved by discussion with the third reviewer if required.

We assessed the included studies using the QUADAS-2 Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies [16]. This was done in duplicate by two reviewers (BDN and TRF), with the final assessment being either the consensus judgment or that agreed following additional discussion. As for the purposes of this review we do not regard each of the QUADAS-2 domains or individual items to have equal importance, rather than creating a summary score we present the assessment separately for each included study, using Review Manager 5.3.

We had intended to carry out meta-analysis of the diagnostic accuracy results to estimate pooled sensitivity and specificity, and present the results in receiver operating characteristic space. After conducting the search it became clear that this would be inappropriate because of the high level of clinical heterogeneity of the included studies. We therefore present the results from each study, including the estimated sensitivity and specificity and 95% confidence intervals [17], stratified by primary tumour type, together with a narrative summary, but do not pool the diagnostic accuracy measures.

3. Results

After removing duplicates and reports that could be excluded by automated title screening, the search identified 2406 studies for assessment (Fig. 1). Of these, 113 required scrutiny of the full text, and nine met the inclusion criteria [18–26]. Of the 104 studies excluded at the final stage of screening, by far the most common reason for exclusion was that the study did not report sufficient information in its results for the full 2 × 2 table of biomarker elevation versus cancer recurrence to be extracted or reconstructed. Often, the reason for this was that biomarker levels were reported only for individuals who had disease recurrence, and not for those who were under surveillance but who had not had a recurrence at the end of the study. Nine studies were excluded at the final stage because they reported baseline or immediately post-surgery marker levels, rather than levels during surveillance, and eight were excluded because they did not consider cancer recurrence. Other reasons for exclusion are shown in Fig. 1.

Among the included studies, the clinical characteristics of patients varied widely. Studies considered patients with a primary diagnosis of seminoma or NSGCT, or used a mixed group (Table 1). Some studies reported a single stage of primary disease; others reported a mixed group. Some studies indicated using an independent reference standard for recurrence, but in several the reference standard was unclear. Most studies reported recurrence on a 'per patient' basis, but some reported 'per sample' – allowing for the possibility of multiple recurrences per patient and treating each negative biomarker result as a 'true negative'. Three studies reported results for LDH; the remainder reported AFP or HCG. Cut-points for defining test positivity varied and in some studies were not reported.

Methodological quality is summarised in accordance with QUADAS-2 guidelines in Fig. 2. There were few applicability concerns, with studies generally using an appropriate patient group, index test and
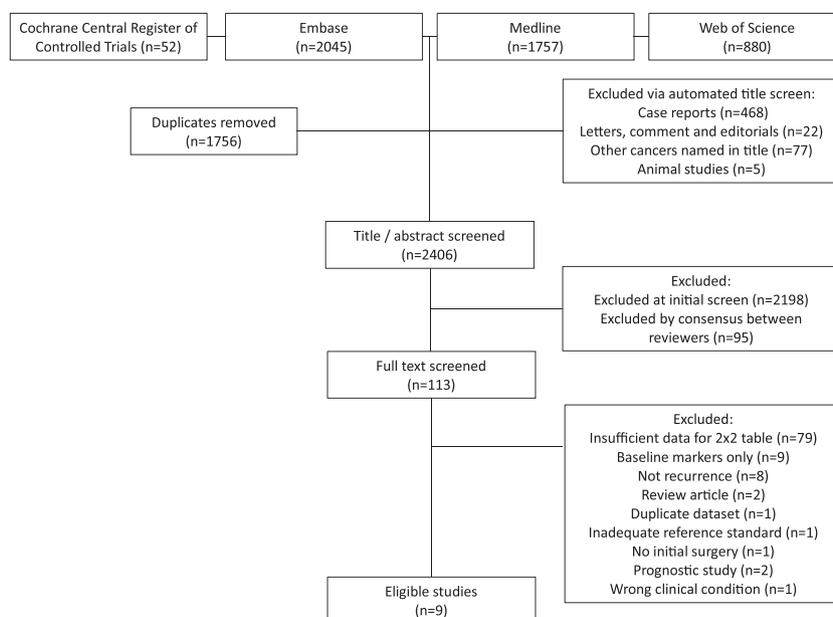
Fig. 1. Study flowchart.

reference standard to match the research question. Risk of bias was high or unclear in at least one domain for each of the included studies. Two major concerns relate to the consistency and timing with which the reference standard (typically, an appropriate imaging modality) was implemented to confirm or refute an elevation in biomarker levels as an indicator of tumour recurrence. Although an inclusion criterion for this review is that the reference standard be performed on both tumour marker positive and tumour marker negative subjects, some studies imply either the possibility of incorporation bias if tumour markers were used directly as part of diagnostic criteria, or that the timing when the reference standard was administered may have differed between marker-positive and marker-negative patients. No studies reported whether the reference standard was interpreted without knowledge of biomarker results. Two studies reported patient drop-out, either for unwillingness to comply with the follow-up protocol [25] or for reasons that were not stated [18]. Two studies used elevation of at least one of the biomarkers as an exclusion criterion without independent verification [25,26].

The high level of heterogeneity precluded a meta-analysis and means it is difficult to draw any firm conclusions about the diagnostic potential of these biomarkers for detecting recurrence. We therefore summarise the findings of the eligible studies narratively. Numerical results regarding diagnostic accuracy estimates are available in Table 1 and Appendix 2.

Three studies reported solely on patients for whom the primary tumour was seminoma [23,25,26]. One found low sensitivity (50%) and moderate specificity (73%) for LDH in the surveillance of a group of 55 patients with Stage I seminoma [25]. Those with elevated AFP were excluded from the cohort. HCG and serum placental alkaline phosphatase (SLAP) were measured pre-operatively but not reported post-operatively. The authors concluded that LDH was not an adequate marker of recurrence in this patient group.

Another study reported HCG for a group of 151 patients with Stage I–IV seminoma and found comparatively low sensitivity (57%) but high specificity (99%), concluding that HCG may have a useful role in monitoring irrespective of the level of this biomarker pre-operatively [23]. Unusually, but in common with De Bruijn et al. [19], this paper also presented longitudinal trends in the monitored HCG level for a subset of patients in the post-operative period, but used a fixed 10 u/l cut-off to define biomarker positivity. HCG was elevated in 7 of the 11 recurrences. Marker-positive recurrences occurred in patients with

stage I–IV disease and in patients both with and without initial marker rise at time of diagnosis. The sample size (11 recurrences among 7 patients) may have been insufficient to incorporate the time trend into a decision strategy.

The final study in this subgroup provided results on a group with Stages I-III seminoma that could not be used for comparative purposes as they were reported at a per-sample level rather than a per-patient level, with an average of approximately four follow-up biomarker measurement visits per patient [26]. HCG and LDH were the markers most frequently used in follow-up. For HCG, they reported very high specificity (98%) and low sensitivity (48%) with 90 positive results among 2790 samples taken, 19 of which were in patients found to have disease recurrence. Results were similar for LDH with 15 recurrences among the 77 elevated results from a total of 2033 samples. Reported specificity was again very high (97%) and sensitivity low (46%). Of 611 placental alkaline phosphatase (PLAP) samples taken, there were 92 positive results but only 6 of these patients had recurrent disease. All 6 patients with disease recurrence in whom PLAP was measured had a positive result. The authors concluded that the combination of these three tumour markers should be used in monitoring for recurrence of seminoma to increase sensitivity, but highlight the high false positive rate, particularly amongst PLAP values in smokers.

Three studies reported results for a mixed group of seminoma and NSGCTs as the primary tumour [18,22,24]. The first of these is a short paper dating from 1978 which, despite showing an apparently promising level of diagnostic accuracy, suffers from several methodological limitations (Fig. 2), including lack of specification of the patient population, index test and reference test, which limit its applicability [18]. Two more recent studies demonstrate much lower sensitivity levels, of 59% and 40%, for HCG and LDH respectively [22,24]. Specificity remains above 90% in both studies. However, at least one of these studies appears likely to suffer from incorporation bias as biomarker elevation explicitly formed part of the criteria for "active disease" [24], while in the other the reference standard was not clearly stated [22]. Misclassification of the timing of recurrence therefore appears possible in both studies.

Of the studies to report solely on patients with NSGCTs, two estimate 100% specificity for AFP, although these use different cut-offs for marker positivity and one combines Stage I-III tumours [19], while the other considers Stage I tumours only [21]. One of these studies estimates sensitivity at 36% [19] and the other at 100% [21], although the

Table 1
Summary of results from included studies reporting data per patient.

Author Date Sample size Median age	Stage Treatment received.	Definition of recurrence. Reference standard	AFP Sens/spec	HCG Sens/spec	LDH Sens/spec	Combinations Sens/spec
SEMINOMA						
Paus 1987 n/N = 7/151 37 years	*Stages I-IV* I-II: Orc, Rt Abdo. III: Orc, Rt Chest Abdo IV: Cx, Rt, Surgery.	Not stated *Not stated*	NR	At 10 u/l 57% [25-84] 99% [95-100]	NR	NR
von Eyben 2002 n/N = 10/55 36 years	*Stage 1* Surveillance	Unclear *Physical examination, CXR, CT and lymphangiography*	NR	NR	At 114 U/L 50% [24-76] 73% [59-84]	NR
SEMINOMA and NSGCTs						
Aiginger 1978 n/N = 21/40 *Not stated*	*Mixed stage.* Unclear.	Abdo lymph nodes / metastases *Unclear*	*Threshold not stated* 86% [65-95] 100% [84-100]	*Threshold not stated* 86% [65-95] 95% [75-99]	NR	*AFP or HCG* 100% [85-1] 95% [75-99]
Lempiäinen 2008 n/N = 22/362 *Not stated*	*Stage I-III* Orc +/-Rt +/- Cx +/- RLND	Not stated *Not stated*	NR	At 2 pmol/l 59% [39-77] 94% [91-96]	NR	NR
Venkitaraman 2007 n/N = 13/449 *Not stated*	*All stages* I: Surveillance II+: "treatment"	Active disease. *Radiological or pathological evidence*	NR	NR	At 1.5x upper limit of normal 40% [20-64] 91% [87-93]	NR
NSGCTs						
De Bruijn 1985 n/N = 22/126 *Not stated*	*Stages I-III* I: Orc and surveillance. II: adjuvant Cx. III – combined Cx.	Active disease after complete remission. *CXR, CT and echography*	At 20 μg/l 36% [20-57] 100% [96-100]	At 2 μg/l 59% [39-77] 98% [93-99]	NR	*AFP or HCG* 86% [67-95] 98% [93-99]
Fosså 1994 n/N = 22/102 *Not stated*	*Stage I (Low-risk)* Orc and surveillance	Unclear *CT Chest and Abdo.*	NR	NR	NR	At AFP 15 μg/l or HCG 10 IU/l 36% [20-57] 95% [88-98]
Kuo 1999 n/N = 1/5 30 years	*Stage I* Radical inguinal Orc	Unclear *Physical examination, CXR and CT*	At 8.5 ng/ml 100% [21-100]	At 10 IU/l 0 % [0-79] 100% [51-100]	NR	NR

Cx: chemotherapy; Rt: Radiotherapy; Orc: Orchidectomy; RPLND: Retroperitoneal Lymph Node Dissection; AFP: Alpha-fetoprotein; CI: confidence interval; CT: computed tomography; CXR: chest X-ray; FU: follow-up; HCG: serum human chorionic gonadotropin; NSGCT: Non-Seminomatous Germ Cell Tumour; n/N: sample size, written as number of recurrences/total number of patients.

Fig. 2. Assessment of methodological quality.

second of these has a wide confidence interval owing to its small sample size (1 recurrence in 5 patients) and suffers from methodological concerns relating to the specification of the reference standard. Sensitivity is increased to 86% in De Bruijn et al. [19], with little adverse effect on specificity, by considering a combined biomarker of AFP and HCG. However, when a similar combination was considered in a purely Stage I group, the sensitivity was much lower (36%) [20].

The available data were not sufficient to make a quantitative assessment of the extent of publication bias. However, the methodological concerns stated above imply that publication bias may substantially affect the quality of the research in this area. Many studies appeared not to pre-specify which biomarkers they would report, and this decision may therefore have been based on observed results, which may in turn have influenced the chance of publication. For example, some studies did not report on more than one biomarker even though it appears likely that multiple biomarkers were measured as part of the routine surveillance strategy for recurrence detection, given these biomarkers are thought to be associated with development of the primary tumour.

4. Discussion

Our review underlines the paucity of evidence underpinning the inclusion of AFP, HCG, and LDH in international guidelines for the surveillance of testicular cancer recurrence [7]. The heterogeneity of the included studies, small sample sizes and the lack of contemporary data preclude any firm conclusions about the accuracy of these biomarkers to detect recurrent testicular cancer. The current literature, therefore, remains inadequate to determine whether there is a subgroup of patients under surveillance for recurrent disease for which a follow-up strategy based on biomarkers (alone, or in combination) could provide an alternative cost-effective means of reducing the intensity of follow-up investigations and hospital episodes included in contemporary multi-modal follow-up.

A general observation is that the specificity of AFP or HCG alone is typically high (90–100%) and, except for some small studies, the sensitivity is much lower, implying that many recurrences would be missed (false negatives) using biomarker assessment alone at the cut-point adopted. The specificity and sensitivity of LDH appeared low in the three studies that reported it [24–26]. The limited data may support the use of biomarkers in combination, as together AFP and HCG appear to have higher sensitivity and lower specificity for seminoma and NSGCT recurrence than when used individually. Marker expression varies between NSGCTs, reflecting the histological heterogeneity of these tumours, so using a combination of biomarkers may increase rates of detection. However, no evidence was found for the use of LDH in combination with AFP and HCG in relation to any tumour type. LDH is a ubiquitous enzyme, and rises are also seen in other sources of tissue damage, such as inflammation [27].

The lack of evidence on the most efficient multimodal follow-up schedule has been highlighted before [9,28]. The resultant variation in follow-up intensity and composition between institutions is often dependent on judgements based on primary tumour stage, histological composition and treatment history [8,9,29,30]. Following concerns over the risks of serial imaging and the financial burden caused by frequent diagnostic evaluations, research into more efficient follow-up is now taking place. Evidence from a randomized controlled trial led to the recommendation to reduce CT frequency from five times to twice per year in years 1–2 of NSGCT follow-up [31,32]. Physical examination has been shown to provide little additional clinical information when used alongside CT [31]. A recent retrospective analysis of two follow-up cohorts led to the recommendation that, in the context of timely cross-sectional imaging, chest radiographs no longer add any value in the routine surveillance of stage I testicular cancer [14]. However, it remains unclear from the literature what is the incremental value in using biomarkers in addition to cross-sectional imaging: studies from the early 1980's are cited to justify the necessity for biomarkers to

avoid false-negative CT examinations [29,33–35]. Although imaging techniques have improved since that time and more recent papers have been published looking at biomarker levels at time of recurrence [36], we retrieved no studies examining the role of biomarkers in relation to modern cross-sectional imaging techniques such as Positron Emission Tomography (PET-CT), that has shown greater sensitivity and specificity for detecting active disease than conventional CT [6].

Some authors have called for routine surveillance of tumour markers to be discontinued given the minimal evidence for detecting early recurrence [37], although it is important to emphasise also that lack of published evidence does not necessarily imply lack of utility as seen from direct clinical experience. Nevertheless, tumour markers continue to be recommended on the basis that they will infrequently identify recurrent disease which might otherwise have been missed, despite the fact the cost implications and false positive rates remain poorly defined [38]. A less intensive, evidence-based, multimodal follow-up strategy in primary care could reduce health care costs, potential harms of over-testing, and the patient anxiety associated with hospital follow-up.

In addition to the development of superior cross-sectional imaging techniques researchers have sought to develop superior biomarkers to overcome the limitations of the three commonly used biomarkers AFP, HCG and LDH. Ideally these would not only have greater sensitivity and specificity but also respond consistently to disease burden and treatment, helping to differentiate between tumour types, such as seminoma and NSGCTs. Markers that have been studied include PLAP, neuron-specific enolase (NSE), TRA-1-60, cell-free circulating DNA, lectin reactive AFP and a range of single and panels of microRNAs [6,10]. A recent prospective study of microRNA miR-371a-3p in 166 patients with GCT found the marker to be elevated in all nine participants with relapsing disease, outperforming the other markers studied such as AFP and bHCG [39]. Whilst these novel markers are promising alternatives, the current evidence-base for these techniques is lacking to support their routine use in detecting testicular cancer recurrence. Comprehensive diagnostic accuracy studies comparing these techniques with the tumour markers included in this review would help determine their future role in the management pathway [6].

By following Cochrane diagnostic test accuracy guidance we included only studies reporting a complete 2 × 2 table. In doing so we excluded a considerable number of studies reporting only sensitivity data. Some of these studies have historically been used to inform testicular cancer follow-up guidelines [7,8,29]. In this regard, our results challenge conventional opinion in this setting.

The several important limitations of studies published in this field must be restated here to improve future evaluations of biomarker performance. Firstly, many studies report biomarker results only in patients who develop recurrence, precluding the estimation of false positive rates and specificity. Incomplete reporting is recognised as an important source of bias and waste in diagnostic accuracy research [40,41]. For example, selection bias leads to overestimation of sensitivity and specificity when consecutive or randomly selected patients are not included [11]. Furthermore, partial verification bias leads to an overestimation of sensitivity with a variable effect on specificity when a non-random selection of patients do not undergo the reference standard. In the studies included in this review, confirmatory testing was commonly only conducted for patients with biomarker elevation. This approach risks non-secreting tumours being misclassified as negatives, inflating specificity estimates. We generally found little analysis of the outcome in those with marker negative results, whilst in general the reported positive marker results were well in excess of the cut-off used. This is a particular concern when studies report 'per sample' data in patients with serial testing and compare this with 'per patient' data. Although a 'per-sample' clearly reflects the way a surveillance strategy is implemented in practice, per-sample information has limited value: the apparent specificity is substantially inflated, as the majority of patients, for whom recurrence does not occur during the follow-up period contribute multiple 'true negative' biomarker measurements. The

resulting very high sensitivity and specificity estimates reported in the Weissbach et al. paper therefore have limited clinical validity [26]. Conversely, a more detailed analysis of false positive results would be required to understand the role of other malignancies or benign conditions that can also cause marker elevation, to avoid over-investigation by establishing thresholds specific to testicular tumour recurrence.

Finally, only three studies reported a single tumour type and stage, whereas in clinical practice follow-up schedules vary based on tumour characteristics and treatment received; we found no evidence to support the frequency of biomarker testing, with some papers reported results for biomarker rise at any point during follow-up, allowing the possibility of delays before confirmatory testing; and we cannot determine the relationship between tumour marker rise and symptom onset, making it impossible to know how much earlier recurrent tumours are detected by current surveillance strategies.

Our results demonstrate a clear opportunity for research to inform the ongoing use of biomarkers for testicular cancer surveillance. Ideally, studies could include existing retrospective cohorts of consecutive patients enrolled onto routine surveillance or prospective cohorts could be set-up for this purpose. Data capture should include details including the timings of symptoms, clinical evaluation, biomarker measurements and imaging in relation to the diagnosis of recurrence. For these data to be included in future diagnostic accuracy reviews, researchers should endeavor to at least report individual 2 × 2 tables for each biomarker and biomarker combination in relation to an appropriate reference standard for testicular cancer recurrence. The publication of individual patient data would also allow future meta-analysis to determine optimal biomarker thresholds for each recurrent tumour type, biomarker performance in relation to other investigative modalities, and the investigation of the value of biomarker trend over and above fixed biomarker thresholds.

5. Conclusions

Having systematically reviewed the available literature, we found surprisingly little evidence to guide optimal testing with biomarkers routinely used during follow-up for testicular cancer recurrence. Without these data, no definitive conclusions can be made about the appropriateness of testing intervals, biomarkers, biomarker cut-points, or patient groups. Data from existing or novel follow-up cohorts should be analysed to inform the most appropriate biomarker testing strategy and how this relates to the performance of modern-day imaging techniques. The quality of the conduct and reporting of diagnostic accuracy studies in this field must vastly improve to facilitate future meta-analysis.

Funding

Declaration of interests

None.

Author contribution statement

BDN, AP, JJ, AVDB and TRF planned and designed the study. NWR assisted with the literature search and identification of papers for inclusion. BDN, NRJ and TRF screened papers for inclusion, extracted and analysed data and drafted the manuscript. All authors contributed to the interpretation of the findings and the revision and final approval of

the manuscript.

Acknowledgements

We thank Sergei Maslau, José Ordóñez-Mena and Thomas E. Jones for help with translations and Robert Watson for helpful discussions.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.canep.2019.01.001.

References

[1] Centers for Disease Control and Prevention, United States Cancer Statistics: An Interactive Cancer Atlas (InCA). https://nccd.cdc.gov/DCPC_INCA. (Accessed 9 January 2018).

[2] Office for National Statistics, Cancer Registration Statistics, England (2017) (Accessed 09 January 2018), https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland.

[3] National Cancer Institute Surveillance, Epidemiology and End Results Program, SEER Cancer Statistics Review (CSR) 1975-2014, (2017) (Accessed 09 January 2018), https://seer.cancer.gov/csr/1975_2014.

[4] Cancer Research UK, Testicular Cancer Survival Statistics, (2017) (Accessed 09 January 2018), http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/testicular-cancer/survival.

[5] International Germ Cell Cancer Collaborative Group, International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers, J. Clin. Oncol. 15 (2) (1997) 594–603.

[6] M.J. Murray, R.A. Huddart, N. Coleman, The present and future of serum diagnostic tests for testicular germ cell tumours, Nat. Rev. Urol. 13 (12) (2016) 715–725.

[7] P. Albers, W. Albrecht, F. Algaba, C. Bokemeyer, G. Cohn-Cedermark, K. Fizazi, A. Horwich, M.P. Laguna, N. Nicolai, J. Oldenburg, Testicular Cancer (Accessed 22 May 2017), http://uroweb.org/guideline/testicular-cancer.

[8] R.J. Groll, P. Warde, M.A.S. Jewett, A comprehensive systematic review of testicular germ cell tumor surveillance, Crit. Rev. Oncol. Hematol. 64 (3) (2007) 182–197.

[9] L.J. Barlow, G.M. Badalato, J.M. McKiernan, Serum tumor markers in the evaluation of male germ cell tumors, Nat. Rev. Urol. 7 (11) (2010) 610–617.

[10] J.C. Milose, C.P. Filson, A.Z. Weizer, K.S. Hafez, J.S. Montgomery, Role of biochemical markers in testicular cancer: diagnosis, staging, and surveillance, Open Access J. Urol. 4 (1) (2012) 1–8.

[11] M.M.G. Leeflang, J.J. Deeks, C. Gatsonis, P.M.M. Bossuyt, Systematic reviews of diagnostic test accuracy, Ann Int Med 149 (12) (2008) 889–897.

[12] N.H. Hanna, L.H. Einhorn, Testicular cancer - discoveries and updates, New. Engl. J. Med. 371 (21) (2014) 2005–2016.

[13] E. Salminen, H. Niiniviita, H. Jarvinen, S. Heinavaara, Cancer death risk related to radiation exposure from Computed Tomography scanning among testicular cancer patients, Anticancer Res. 37 (2) (2017) 831–834.

[14] H. De La Pena, A. Sharma, C. Glicksman, J. Joseph, M. Subesinghe, Z. Traill, C. Verrill, M. Sullivan, J. Redgwell, E. Bataillard, E. Pintus, N. Dallas, A. Gogbashian, M. Tuthill, A. Protheroe, M. Hall, No longer any role for routine follow-up chest x-rays in men with stage I germ cell cancer, Eur. J. Cancer 84 (2017) 354–359.

[15] B. Nicholson, N. Jones, A. van den Bruel, A. Protheroe, J. Joseph, T. Fanshawe, Biomarkers for the Detection of Testicular Cancer Recurrence, CRD42017074683 PROSPERO, 2017 (Accessed 09 January 2018), http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017074683.

[16] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M.M. Leeflang, J.A. Sterne, P.M. Bossuyt, QUADAS-2 Group, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (18) (2011) 529–536.

[17] S. Dorai-Raj, Binom: binomial Confidence Intervals for Several Parameterizations. R Package Version 1.1-1, (2014) https://CRAN.R-project.org/package=binom.

[18] P. Aiginger, H. Kolbe, J. Kühböck, J. Spona, Diagnostische Fortschritte bei malignen Hodentumoren [Diagnostic progress in testicular carcinoma], Acta Med. Austriaca 5 (4–5) (1978) 159–161.

[19] H.W.A. De Bruijn, D.T. Sleijfer, H. Schraffordt Koops, A.J.H. Suurmeijer, J. Marrink, T. Ockhuizen, Significance of human chorionic gonadotrophin, alpha-fetoprotein, and pregnancy-specific beta-1-glycoprotein in the detection of tumor relapse and partial remission in 126 patients with nonseminomatous testicular germ cell tumors, Cancer 55 (4) (1985) 829–835.

[20] S.D. Fosså, A.B. Jacobsen, N. Aass, A. Heilo, A.E. Stenwig, O. Kummen, N.B. Johannessen, G. Waaler, P. Øgreid, L. Borge, T. Urnes, T. Bjerklund-Johansen, How safe is surveillance in patients with histologically low-risk non-seminomatous testicular cancer in a geographically extended country with limited computerised tomographic resources? Br. J. Cancer 70 (6) (1994) 1156–1160.

[21] J.-Y. Kuo, T. Chin, Y.-L. Hsieh, A.T.L. Lin, Y.-H. Chang, C. Wei, K.-K. Chen, L.S. Chang, Observations after orchiectomy in clinical stage I nonseminomatous germ cell tumors of the testis, Zhonghua Yi Xue Za Zhi (Taipei) 62 (6) (1999) 356–361.

[22] A. Lempiäinen, U.-H. Stenman, C. Blomqvist, K. Hotakainen, Free ß-subunit of human chorionic gonadotropin in serum is a diagnostically sensitive marker of seminomatous testicular cancer, Clin. Chem. 54 (11) (2008) 1840–1843.

[23] E. Paus, S.D. Fosså, T. Risberg, K. Nustad, The diagnostic value of human chorionic gonadotrophin in patients with testicular seminoma, Br. J. Urol. 59 (6) (1987) 572–577.

[24] R. Venkitaraman, B. Johnson, R.A. Huddart, C.C. Parker, A. Horwich, D.P. Dearnaley, The utility of lactate dehydrogenase in the follow-up of testicular germ cell tumours, BJU Int. 100 (1) (2007) 30–32.

[25] F.E. von Eyben, E.L. Madsen, O. Blaabjerg, P.H. Petersen, G.K. Jacobsen, L. Specht, B.N. Pedersen, H. von der Maase, Serum lactate dehydrogenase isoenzyme 1 in patients with seminoma stage I followed with surveillance, Acta Oncol. 41 (1) (2002) 77–83.

[26] L. Weissbach, R. Bussar-Maatz, K. Mann, The value of tumor markers in testicular seminomas - Results of a prospective multicenter study, Eur. Urol. 32 (1) (1997) 16–22.

[27] D. Weatherby, S. Ferguson, Blood Chemistry and CBC Analysis, Bear Mountain Publishing, Jacksonville, OR, 2004.

[28] T.D. Gilligan, J. Seidenfeld, E.M. Basch, L.H. Einhorn, T. Fancher, D.C. Smith, A.J. Stephenson, D.J. Vaughn, R. Cosby, D.F. Hayes, O. American Society of Clinical, American Society of Clinical Oncology Clinical Practice Guideline on uses of serum tumor markers in adult males with germ cell tumors, J. Clin. Oncol. 28 (20) (2010) 3388–3404.

[29] N.J. van As, D.C. Gilbert, J. Money-Kyrle, D. Bloomfield, S. Beesley, D.P. Dearnaley, A. Horwich, R.A. Huddart, Evidence-based pragmatic guidelines for the follow-up of testicular cancer: optimising the detection of relapse, Br. J. Cancer 98 (12) (2008) 1894–1902.

[30] G.V. Kondagunta, J. Sheinfeld, R.J. Motzer, Recommendations of follow-up after treatment of germ cell tumors, Semin. Oncol. 30 (3) (2003) 382–389.

[31] G.J. Rustin, G.M. Mead, S.P. Stenning, P.A. Vasey, N. Aass, R.A. Huddart, M.P. Sokal, J.K. Joffe, S.J. Harland, S.J. Kirk, Randomized trial of two or five computed tomography scans in the surveillance of patients with stage I non-seminomatous germ cell tumors of the testis: medical research council trial TE08, ISRCTN56475197—the National Cancer Research Institute Testis Cancer Clinical Studies Group, J. Clin. Oncol. 25 (11) (2007) 1310–1315.

[32] K. Oechsle, A. Lorch, F. Honecker, C. Kollmannsberger, J.T. Hartmann, I. Boehlke, J. Beyer, C. Bokemeyer, Patterns of relapse after chemotherapy in patients with high-risk non-seminomatous germ cell tumor, Oncology 78 (1) (2010) 47–53.

[33] J.L. Thomas, M.E. Bernardino, R.B. Bracken, Staging of testicular carcinoma: comparison of CT and lymphangiography, Am. J. Roentgenol. 137 (5) (1981) 991–996.

[34] J.P. Richie, M.B. Garnick, H. Finberg, Computerized Tomography: how accurate for abdominal staging of testis tumors? J. Urol. 127 (4) (1982) 715–717.

[35] R.G. Rowland, D. Weisman, S.D. Williams, L.H. Einhorn, E.C. Klatte, J.P. Donohue, Accuracy of preoperative staging in stages A and B of nonseminatous germ cell testis tumors, J. Urol. 127 (4) (1982) 718–720.

[36] J. Trigo, J. Tabernero, L. Paz-Ares, J. Garcia-Llano, J. Mora, P. Lianes, E. Esteban, R. Salazar, J. Lopez-Lopez, H. Cortes-Funes, Tumor markers at the time of recurrence in patients with germ cell tumors, Cancer 88 (1) (2000) 162–168.

[37] D. Vesprini, P. Chung, S. Tolan, M. Gospodarowicz, M. Jewett, M. O'Malley, J. Sweet, M. Moore, T. Panzarella, J. Sturgeon, L. Sugar, L. Anson-Cartwright, P. Warde, Utility of serum tumor markers during surveillance for stage I seminoma, Cancer 118 (21) (2012) 5245–5250.

[38] M.S. Mortensen, J. Lauritsen, M.G. Gundgaard, M. Agerbaek, N.V. Holm, I.J. Christensen, H. von der Maase, G. Daugaard, A nationwide cohort study of stage I seminoma patients followed on a surveillance program, Eur. Urol. 66 (6) (2014) 1172–1178.

[39] K.P. Dieckmann, A. Radtke, M. Spiekermann, T. Balks, C. Matthies, P. Becker, C. Ruf, C. Oing, K. Oechsle, C. Bokemeyer, J. Hammel, S. Melchior, W. Wosniok, G. Belge, Serum levels of MicroRNA miR-371a-3p: a sensitive and specific new biomarker for germ cell tumours, Eur. Urol. 71 (2) (2017) 213–220.

[40] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L. Irwig, J.G. Lijmer, D. Moher, D. Rennie, H.C. de Vet, H.Y. Kressel, N. Rifai, R.M. Golub, D.G. Altman, L. Hooft, D.A. Korevaar, J.F. Cohen, STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies, BMJ 351 (2015) h5527 (Clinical research ed.).

[41] P. Glasziou, D.G. Altman, P. Bossuyt, I. Boutron, M. Clarke, S. Julious, S. Michie, D. Moher, E. Wager, Reducing waste from incomplete or unusable reports of biomedical research, Lancet 383 (9913) (2014) 267–276.