Sarah J Price, Elizabeth A Shephard, Sally A Stapley, Kevin Barraclough and William T Hamilton

Non-visible versus visible haematuria and bladder cancer risk:

a study of electronic records in primary care

Abstract

Background
Diagnosis of bladder cancer relies on investigation of symptoms presented to primary care, notably visible haematuria. The importance of non-visible haematuria has never been estimated.

Aim
To estimate the risk of bladder cancer with non-visible haematuria.

Design and setting
A case–control study using UK electronic primary care medical records, including uncoded data to supplement coded records.

Method
A total of 4915 patients (aged ≥40 years) diagnosed with bladder cancer between January 2000 and December 2009 were selected from the Clinical Practice Research Datalink and matched to 21 718 controls for age, sex, and practice. Variables for visible and non-visible haematuria were derived from coded and uncoded data. Analyses used multivariable conditional logistic regression, followed by estimation of positive predictive values (PPVs) for bladder cancer using Bayes' theorem.

Results
Non-visible haematuria (coded/uncoded data) was independently associated with bladder cancer: odds ratio (OR) 20 (95% confidence interval [CI] =12 to 33). The PPV of non-visible haematuria was 1.6% (95% CI = 1.2 to 2.1) in those aged ≥60 years and 0.8% (95% CI = 0.1 to 5.6) in 40–59-year-olds. The PPV of visible haematuria was 2.8% (95% CI = 2.5 to 3.1) and 1.2% (95% CI = 0.6 to 2.3) for the same age groups respectively, lower than those calculated using coded data alone. The proportion of records of visible haematuria in coded, rather than uncoded, format was higher in cases than in controls (P<0.002, χ^2 test). There was no evidence for such differential recording of non-visible haematuria by case/control status (P = 0.78), although, overall, the uncoded format was preferred (P<0.001).

Conclusion
Both non-visible and visible haematuria are associated with bladder cancer, although the visible form confers nearly twice the risk of cancer compared with the non-visible form. GPs' style of record keeping varies by symptom and possible diagnosis.

Keywords
bladder cancer; diagnosis; haematuria; primary care.

INTRODUCTION

Bladder cancer is common, with more than 10 000 new diagnoses in the UK annually[1] and in excess of 5000 deaths.[2] Comparison of UK data with the European Union average shows that the age-standardised incidence rate for men is much lower in the UK (14.8 versus 29.1 per 100 000), whereas the rates for women are much the same (4.5 versus 6.1 per 100 000).[3] Mortality is strongly related to the cancer stage at diagnosis, with tumours that have invaded muscle resulting in a poor prognosis. Bladder cancer is more common with advancing age, and in men, and is strongly related to cigarette smoking. No screening is available, so diagnosis relies on symptomatic presentation, generally to primary care.[4]

Dysuria, abdominal pain, and loss of appetite or weight are all associated with bladder cancer,[4,5] but the most common and highest-risk symptom in primary care is haematuria.[4–8] Haematuria may be recognised by the patient; so-called 'frank', 'visible', or 'macroscopic' haematuria.[9] Alternatively, it may be detected only on examination of urine, most commonly as a positive urinalysis test for blood (chemical dipstick), or as more than a set number of red cells per high-power field on microscopy, or in a Coulter counter.[10] Under these circumstances, it is called 'non-visible', 'invisible', or 'microscopic' haematuria. Isolated non-visible haematuria is defined as three or more red blood cells per high-power field in the absence of infection or proteinuria.[9,11] What is considered 'normal' varies greatly as healthy people lose around a million red blood cells in their urine daily, equating to around one cell per high-power field. Chemical dipsticks give a 'negative' finding with this level of haematuria. A recent systematic review concluded that dipsticks are a reasonable method to use in isolation (positive likelihood ratio 5.99, 95% confidence interval (CI) = 4.04 to 8.89, negative likelihood ratio 0.21, 95% CI = 0.17 to 0.26). Therefore, for this and other reasons related to cost and practicality, confirmation of a positive dipstick result by urinary microscopy in primary care is no longer recommended.[10,11]

The background rate of asymptomatic non-visible haematuria in UK males is around 2.5%, increasing to 22% in those aged >60 years.[10] Other reports also show higher prevalence with increasing age.[11,12] A screening study of annual urinalysis in 1000 healthy Israeli military male recruits, aged 18–33 years at the beginning of the 15-year study, reported 39% with dipstick-positive urinalysis for haematuria at least once and 16% on two or more occasions.[13]

Visible haematuria is accepted as an important pointer to urinary tract cancer (with bladder cancer being the most common), warranting urgent urological investigation, usually by ultrasound and cystoscopy. UK referral guidelines published in 2005 by the National Institute for Health and Clinical Excellence (NICE, now the

SJ Price, BSc, PhD student; **EA Shephard**, PhD, associate research fellow; **SA Stapley**, PhD, associate research fellow; **WT Hamilton**, MD, FRCP, FRCGP, professor of primary care diagnostics, University of Exeter Medical School, Exeter. **K Barraclough**, MA, FRCP, FRCGP, LLB, GP, Hoyland House General Practice, Painswick, Gloucestershire.

Address for correspondence
William T Hamilton, University of Exeter Medical School, College House, St Luke's Campus, Magdalen Road, Exeter, EX1 2LU.

E-mail: w.hamilton@exeter.ac.uk

Submitted: 26 February 2014; **Editor's response:** 1 April 2014; **final acceptance:** 11 April 2014.

This is the full-length article (published online 1 Sep 2014) of an abridged version published in print. Cite this article as: **Br J Gen Pract 2014; DOI: 10.3399/bjgp14X681409.**

How this fits in

Visible haematuria is well recognised as a risk marker for bladder cancer. This study shows that non-visible haematuria is also a risk marker, although it is approximately half as likely as visible haematuria to be indicative of the disease. Using previously 'hidden' data in electronic records, namely uncoded 'free-text' notes recorded by GPs, this study estimated the above association, which had previously been impossible in analyses of coded data alone. The analysis of coded and uncoded data revealed differential patterns of recording, suggesting that GPs record 'red-flag' symptoms in the most-visible format.

National Institute for Health and Care Excellence) are currently being updated, but recommend referral for patients of any age with visible haematuria, after identification and treatment of any urinary tract infection.[14] For unexplained non-visible haematuria, the recommendation for patients aged <50 years is non-urgent referral; urgent referral is recommended for those with the same symptoms who are aged ≥50 years.[14] National guidance focusing on chronic kidney disease makes similar recommendations for the prompt investigation of non-visible haematuria.[11]

Large electronic primary care databases, for example the Clinical Practice Research Datalink (CPRD), are increasingly used in epidemiological studies.[15] Such databases store information in either coded or uncoded format, with the coded data alone being used in almost all studies to date. Coded CPRD data are stored as medcodes (a CPRD-specific term), of which there are over 100 000, allowing detailed description of primary care events. Data in uncoded format include 'free-text' notes added by GPs to supplement coded entries. The uncoded portion of the CPRD is not routinely studied for several reasons, including its analytical complexity and to protect patient anonymity.

Two studies have added uncoded data to standard methods: one showed that 10% of suicides were identifiable in uncoded format only;[16] the other that 29% of patients with rheumatoid arthritis have an uncoded record suggestive of the diagnosis in advance of the first coded entry for the disease.[17] Three large primary care database studies have examined visible haematuria and urological cancer.[4,5,8] Only one of these also considered non-visible haematuria; however, this was omitted from the main analyses, being deemed too infrequent to provide reliable estimates.[4]

For visible haematuria, positive predictive values (PPVs) ranged from 3.4% (females aged >15 years), 3.9% (either sex, aged >60 years)[4] to 7.5% (males aged >15 years)[8] and 12.5% (male, heavy smoker, aged 70 years).[5] A fourth, smaller study of urinary tract cancer reported a PPV for visible haematuria of 10.3%.[6] For non-visible haematuria, a US cohort study of urological evaluation of secondary care patients with non-visible haematuria found urinary tract cancers in 2.1% of the patients.[18] Secondary care reports suggest the risk of cancer is much lower from non-visible than visible haematuria;[7,19] whether these results translate to the lower-incidence population of primary care is unknown.

This study uses the full potential of CPRD data by analysing both coded and uncoded data to estimate bladder cancer risk in primary care patients with non-visible haematuria.

METHOD

This study extends a previous study of the clinical features of bladder cancer in primary care[4] by adding uncoded data. A brief summary of the first study is given here, along with the methods used for extension.

The studies both have a matched case–control design using the UK's General Practice Research Database (now the CPRD). Cases (n = 4915) with bladder cancer codes reported between January 2000 and December 2009 inclusive were matched with up to five controls (n = 21 718) on sex, age, and general practice. All were ≥40 years old, with at least 1 year of data before diagnosis. Features identified by univariable analysis as having an independent association with bladder cancer were entered into multivariable analysis and PPVs for each were estimated using Bayes' theorem: prior odds × likelihood ratio = posterior odds, where prior odds are the age-specific national incidence for bladder cancer (2008), expressed as odds. An erratum was published in March 2014, changing some PPVs as a minor arithmetical error had affected the haematuria results. In this study, the corrected PPVs have been used throughout to make comparisons.

Uncoded data

Extracts of uncoded records of cases and controls were requested for the year immediately preceding the patient's diagnosis of bladder cancer. Each extract

Table 1. Visible and non-visible haematuria in cases and controls, according to type of record

	Cases, *n* = 4915 (%)			Controls, *n* = 21 718 (%)		
Type of haematuria	Coded	Uncoded	Coded or uncoded	Coded	Uncoded	Coded or uncoded
Visible	2595 (52.8)	1991 (40.5)	3147 (64.0)	196 (0.9)	206 (0.9)	336 (1.5)
Non-visible	127 (2.6)	216 (4.4)	313 (6.4)	26 (0.1)	41 (0.2)	60 (0.3)

Records of haematuria that were in both coded and uncoded format were assumed to refer to a single occurrence of the symptom and counted only once; as such, the total in coded or uncoded format is not the sum of the coded plus uncoded records.

(*n* = 4666) contained the expression 'haematuria' or 'blood in urine', plus up to three words either side to give context, such as 'no haematuria'. These uncoded entries were categorised according to:

- the type of haematuria described (visible or non-visible); and
- whether haematuria was present or absent at the time of consultation.

The above classification was made using an algorithm that identified descriptors for negation (for example, 'haematuria absent') and non-visible haematuria (for example, 'micro' and 'dipstick'), as well as terms that indicated uncertainty (for example, 'if any' and 'ago'). Extracts flagged as 'uncertain' were then classified manually, in consultation with a GP. The algorithm's validity was assessed by comparing its output with that of a reference standard for visible haematuria. This reference standard was constructed from a random sample of 100 observations by two independent raters (both GPs) using the consensus method.[20] Some uncoded data were difficult to interpret, so the raters had the option of choosing the category 'unclear'; only 2% of extracts were classified as such by both raters. Full disagreement was defined as 'haematuria absent' versus 'haematuria present', whereas partial disagreement occurred when one of the raters opted to classify an extract as unclear. To err on the side of caution, the category 'haematuria absent' was assigned wherever there was partial disagreement and to the 2% of extracts that could not be interpreted at all.

For the reference standard, observer inter-rater agreement was good to substantial (weighted κ = 0.7, 95% confidence interval [CI] = 0.5 to 0.9).[21,22] Full disagreement occurred in only 4% of extracts, with partial disagreement in 12%. The main source of disagreement was whether the symptom was historic or an ongoing concern, so the algorithm was optimised to identify terms indicating this. All extracts were classified initially by the algorithm; after manual review, only 59 of the 4666 (1.3%) extracts were considered insufficiently clear and were categorised as 'haematuria absent'. The algorithm performed with 96% sensitivity and 45% specificity. Agreement between the algorithm and the reference standard was good to substantial (weighted κ 0.8, 95% CI = 0.7 to 0.9).[21,22] A check of all complete words identified a mis-spelling rate of only 0.64% and no US spellings.

After this process, new variables for visible and non-visible haematuria were created: patients were deemed to have the symptom if it was present in either coded or uncoded format. The multivariable analyses of the original study were replicated. Clinically plausible interaction terms were sought in the new model to ensure that it fully explained any modification of the effect of one independent variable by another. Data analysis was conducted using Stata (version 12).

RESULTS

The CPRD provided 29 033 patients (4935 cases, 24 098 controls). Application of exclusion criteria resulted in 26 633 patients (4915 cases, 21 718 controls). Details of patient demographics and exclusion criteria are given in the original article.[4] Adding the uncoded data increased the numbers of patients with both forms of haematuria considerably (Table 1). The proportion of records of visible haematuria in coded, rather than uncoded, format was higher in cases than in controls (P<0.002, χ^2 test). There was no evidence for such differential recording of non-visible haematuria by case/control status (P = 0.78), although, overall, the uncoded format was preferred (P<0.001).

In the new analysis, all the original non-haematuria features retained their association with cancer ($P \leq 0.001$ for all variables) with largely unchanged ORs.[4] The OR for visible haematuria reduced from 34 (95% CI = 29 to 41) to 26 (95% CI = 22 to 30) per attendance; the OR for non-visible haematuria was 20 (95% CI = 12 to 33). Two interaction terms were found, both antagonistic: between non-visible and visible haematuria (interaction term 0.04, P<0.001) and between non-visible haematuria and urinary tract infection (0.07, P<0.001).

Positive predictive values

The PPV of visible haematuria for bladder

Table 2. PPVs for clinical features of bladder cancer in patients aged ≥60 years, in combination with visible or non-visible haematuria

	PPV (95% CI) for feature when presented with:	
Clinical feature	**Visible haematuria**	**Non-visible haematuria**
Dysuria	4.1 (2.6 to 6.3)	4.5
Abdominal pain	2.3 (1.5 to 3.5)	1.7 (0.6 to 4.2)
Constipation	2.2 (1.5 to 3.4)	2.0
Urinary tract infection	2.2 (1.8 to 2.8)	1.4 (0.8 to 2.4)
Raised white cell count	3.7 (2.1 to 6.3)	3.9
Raised inflammatory markers	3.3 (2.0 to 5.4)	1.3
Raised creatinine	2.9 (2.1 to 3.9)	1.1 (0.6 to 2.2)

PPV = positive predictive value. 95% CIs have not been calculated when any cell in the 2×2 table was <5 (invariably because too few controls had both features).

cancer was considerably lower in the new analysis: 2.8% (95% CI = 2.5 to 3.1) instead of 3.9% (95% CI = 3.4 to 4.4) for those aged ≥60 years. Similarly, in the group aged 40–59 years, the PPV for visible haematuria fell from 3.1% (95% CI = 1.0 to 9.8) to 1.2% (95% CI = 0.6 to 2.3).

The PPV for non-visible haematuria was 1.6% (95% CI = 1.2 to 2.1) in those aged ≥60 years and 0.8% (95% CI = 0.1 to 5.6) in 40–59-year-olds, although the latter figure is based on small numbers.

Table 2 shows the PPVs for other clinical features of bladder cancer from the original article[4] in combination with visible or non-visible haematuria in those aged ≥60 years.

DISCUSSION

Summary

To the authors' knowledge, this study is the first to report the risk of bladder cancer with non-visible haematuria in primary care. In those aged ≥60 years the risk was 1.6% (95% CI = 1.2 to 2.1). This is almost half that for patients with visible haematuria, but still higher than that in patients presenting with any of the single features of bladder cancer identified in a previous study (data not shown).[4] Including uncoded data in the analysis enabled an estimation of risk for non-visible haematuria and improved the analysis of existing variables.

This is particularly relevant as GPs unexpectedly favoured the coding of visible haematuria in patients who later received a cancer diagnosis compared with those who did not.

Strengths and limitations

The main strengths of this study are that it was set in primary care with data retrieved from a large database; data are not affected by recall bias as they were recorded before cancer was identified. Another advantage is the use of uncoded data, which allowed for the identification of sufficient records of non-visible haematuria for reliable estimates to be obtained.

This study shares the limitations that arise from the use of computerised data, the predominant one being a reliance on accurate symptom reporting and recording. Use of uncoded data reduces the latter of these, but brings problems of its own. The CPRD methods will have missed any uncoded records containing incorrect or US spellings, such as 'hematuria', although no US spelling was found elsewhere and there was a very low typographical error rate. Even so, the numbers of patients with visible or non-visible haematuria are probably slightly underestimated.

Conversion of text extracts to categorical variables will inevitably incur some error and misclassification. Including manual review of extracts flagged as potentially unclear minimised this, leaving only 1.3% of extracts as 'unclear'. Erring on the side of caution, these were categorised as 'haematuria absent', with the consequence that the PPVs may be very slightly underestimated.

Finally, the thoroughness of methods used to extract the uncoded data is not documented by the CPRD but the researchers have no reason to consider it imperfect.

Comparison with existing literature

There are no direct comparison studies. Secondary care studies are an inadequate proxy for primary care owing to selection bias, as not all primary care patients with haematuria receive full investigation.[7] There may be less selection bias in US secondary care studies, where open access to specialists is the norm. Indeed, the results of this study are broadly in line with those of a study of patients with asymptomatic microscopic haematuria in that setting, which found an overall urinary tract cancer rate of 2.1%.[18]

The annual incidence of non-visible haematuria in controls in this current study was 0.3%. This is lower than the reported prevalence in the UK adult male population of 2.5%,[10] and may be explained, in part, by few UK patients having routine urinalysis.

Implications for practice

Causes of non-visible haematuria can be grouped as glomerular (such as immunoglobulin A nephropathy) or non-glomerular (for example, bladder cancer or urinary tract infection).[12] This

explains current NICE guidance, which recommends non-urgent referral for those aged <50 years presenting with non-visible haematuria without a glomerular cause, but urgent referral for all those aged ≥50 years with unexplained non-visible haematuria.[14] The PPVs of 0.8% and 1.6% in 40–59-year-olds and those aged ≥60 years, respectively, support looking for non-malignant diagnoses initially, with urinary tract infection being the likely main cause. However, recent reports suggest that patients would opt for investigation at much lower levels than those recommended by NICE, with almost 90% choosing investigation at a risk of 1%.[23]

This correct study excluded the younger population in whom glomerular causes of non-visible haematuria are likely to predominate; however, this age group was chosen to reflect the fact that bladder cancer is very rare in those aged <40 years.[9]

Not surprisingly, the association between non-visible haematuria and bladder cancer was modified by the presence of visible haematuria (interaction term 0.04), as these symptoms are along a continuum. Nevertheless, even after accounting for interaction, the association between non-visible haematuria and bladder cancer was clear and the results support the decision to refer patients presenting with this symptom, arguably including all those aged ≥40 years rather than ≥50 years as is currently the case. There is less equivocation about visible haematuria; this study reports that the PPVs for this symptom have reduced a little, but arguably support a policy of investigation of those aged ≥40 years.

The results of this study also have implications for those conducted in large electronic datasets that analyse coded data only. The preferential use of a coded format for visible haematuria is important; if this differential coding is also used for other 'red-flag' symptoms of other cancers, then PPV estimates may be artificially high. This finding should be tested in other cancers and their key symptoms. For non-visible haematuria (generally perceived to be less indicative of bladder cancer compared with visible haematuria) GPs used the uncoded format in preference to the coded one to the extent that there were insufficient coded records to allow it to be analysed at all.

GPs' style of record keeping varies by symptom and by possible diagnosis. Both non-visible and visible haematuria are associated with bladder cancer, with the latter having higher risk. Referral is currently recommended and now has an underpinning evidence base. Researchers should be aware that they risk missing important information if they omit uncoded data in studies using CPRD data, potentially introducing errors that may inflate or reduce estimates.

Funding

Sarah J Price is funded by a University of Exeter PhD studentship but there was no specific funding for this study.

Ethical approval

Independent Scientific Advisory Committee: protocol 09-110.

Provenance

Freely submitted; externally peer reviewed.

Competing interests

William T Hamilton is the clinical lead for the ongoing revision of National Institute for Health and Care Excellence (NICE) guidance on investigation of suspected cancer; this article was written in a personal capacity and should not be taken to represent the view of NICE or the Guideline Development Group. The other authors have no competing interests.

Discuss this article

Contribute and read comments about this article: **bjgp.org.uk/letters**

REFERENCES

1. Office for National Statistics. *Cancer Registration Statistics, England, 2011.* http://www.ons.gov.uk/ons/dcp171778_315795.pdf (accessed 5 Aug 2014).
2. Cancer Research UK. *Cancerstats.* 2011. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bladder/?script=true (accessed 5 Aug 2014).
3. International Agency for Research on Cancer. *Bladder Cancer. Estimated incidence, mortality & prevalence in men, 2012.* http://eco.iarc.fr/eucan/Cancer.aspx?Cancer=32 (accessed 5 Aug 2014).
4. Shephard EA, Stapley S, Neal RD, *et al.* Clinical features of bladder cancer in primary care. *Br J Gen Pract* 2012; DOI:10.3399/bjgp12X654560.
5. Hippisley-Cox J, Coupland C. Identifying patients with suspected renal tract cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2012; DOI:10.3399/bjgp12X636074.
6. Bruyninckx R, Buntinx F, Aertgeerts B, Van Casteren V. The diagnostic value of macroscopic haematuria for the diagnosis of urological cancer in general practice. *Br J Gen Pract* 2003; **53(486):** 31–35.
7. Summerton N, Mann S, Rigby AS, *et al.* Patients with new onset haematuria: assessing the discriminant value of clinical information in relation to urological malignancies. *Br J Gen Pract* 2002; **52(477):** 284–289.
8. Jones R, Latinovic R, Charlton J, Gulliford MC. Alarm symptoms in early diagnosis of cancer in primary care: cohort study using General Practice Research Database. *BMJ* 2007; **334(7602):** 1040.
9. Rodgers M, Nixon J, Hempel S, *et al.* Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation. *Health Technol Assess* 2006; **10(18):** iii–iv, xi–259.
10. Kelly JD, Fawcett DP, Goldberg LC. Assessment and management of non-visible haematuria in primary care. *BMJ* 2009; **338:** a3021.
11. National Collaborating Centre for Chronic Conditions. *Chronic kidney disease: national clinical guideline for early identification and management in adults in primary and secondary care.* London: Royal College of Physicians (UK), 2008.
12. Cohen RA, Brown RS. Clinical practice. Microscopic hematuria. *N Engl J Med* 2003; **348(23):** 2330–2338.
13. Froom P, Ribak J, Benbassat J. Significance of microhaematuria in young adults. *Br Med J (Clin Res Ed)* 1984; **288(6410):** 20–22.
14. National Institute for Health and Clinical Excellence. *Referral guidelines for suspected cancer.* London: NICE, 2005.
15. Shephard E, Stapley S, Hamilton W. The use of electronic databases in primary care research. *Fam Pract* 2011; **28(4):** 352–354.
16. Thomas KH, Davies N, Metcalfe C, *et al.* Validation of suicide and self-harm records in the Clinical Practice Research Datalink. *Br J Clin Pharmacol* 2013; **76(1):** 145–157.
17. Ford E, Nicholson A, Koeling R, *et al.* Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? *BMC Med Res Methodol* 2013; **13:** 105.
18. Loo RK, Lieberman SF, Slezak JM, *et al.* Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria. *Mayo Clin Proc* 2013; **88(2):** 129–138.
19. Edwards TJ, Dickinson AJ, Gosling J, *et al.* Patient-specific risk of undetected malignant disease after investigation for haematuria, based on a 4-year follow-up. *BJU Int* 2011; **107(2):** 247–252.
20. Rutjes AW, Reitsma JB, Coomarasamy A, *et al.* Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007; **11(50):** iii, ix–51.
21. McGinn T, Wyer PC, Newman TB, *et al.* Tips for learners of evidence-based medicine: 3. measures of observer variability (kappa statistic). *CMAJ* 2004; **171(11):** 1369–1373.
22. Altman DG. *Practical statistics for medical research.* London: Chapman and Hall, 1991.
23. Banks J, Hollinghurst S, Bigwood L, *et al.* Preferences for cancer investigation: a vignette-based study of primary-care attendees. *Lancet Oncol* 2014; **15(2):** 232–240.