The Aarhus Statement: Improving design and reporting of studies on early cancer diagnosis

A paper for submission to:

British Journal of Cancer

Weller D[1], Vedsted P[2], Rubin G[3], Walter FM[4], Emery J[5], Scott S[6], Campbell C[1], Andersen RS[2], Hamilton W[7], Olesen F[2], Rose P[8], Nafees S[9], van Rijswijk E[10], Hiom S[11], Muth C[12], Beyer M[12], Neal RD[9]

1. Centre for Population Health Sciences, University of Edinburgh
2. Research Unit for General Practice, Aarhus University
3. School of Medicine and Health, Durham University
4. General Practice & Primary Care Research Unit, University of Cambridge
5. School of Primary, Aboriginal and Rural Health Care, University of Western Australia
6. Unit of Social & Behavioural Sciences, Dental Institute, King's College London
7. Primary Care Diagnostics, Peninsula College of Medicine & Dentistry
8. Department of Primary Health Care, University of Oxford,
9. North Wales Centre for Primary Care Research, Bangor University
10. University Medical College, St. Radboud
11. Cancer Research UK, London
12. Institute for General Practice, Johann Wolfgang Goethe University

Corresponding Author - Professor David Weller: david.weller@ed.ac.uk

Running head: The Aarhus Check-list – Early cancer research guidance

Word count (text only):   2824

Background

Early diagnosis of symptomatic cancer is considered to be central to the achievement of better outcomes.[1] There are national initiatives in many countries with the intention of achieving earlier presentation with symptoms, and earlier diagnosis through more effective diagnostic routes. For example, the English National Awareness and Early Diagnosis Initiative (NAEDI)[1] is a programme of work intended to achieve earlier presentation of people with symptoms, to optimise clinical practice and to improve GP access to diagnostics. Similar important initiatives are being undertaken in a growing number of jurisdictions.[2]

Variations in cancer survival between countries of comparable affluence and health systems have prompted international efforts to better understand these differences, often with a focus on the role of earlier diagnosis.[3] The scope for interventions to improve early diagnosis is wide, and is receiving significant investment by national governments. Several of these have acknowledged the need for complementary programmes of research in order to gain better understanding of patient and clinician behaviour and to determine the effectiveness of the interventions they are funding.[4,5]

Research in this area is, however, complex; pathways to cancer diagnosis are often not straightforward and rarely linear.[6,7] The health system context in which research is carried out has a major influence on individual stages of these pathways. Moreover, the published research to date is characterised by a wide range of methodological approaches, often poorly or imprecisely described and with little theoretical basis. A wide range of methods have been used to measure time points and intervals, often dictated by the available data and resources.[8] As a result, research findings are frequently difficult to interpret and typically not comparable beyond individual countries.

Furthermore, a number of different theoretical and methodological perspectives have been applied to the definition and measurement of time points and intervals in early diagnosis research, and these perspectives are often not explicitly delineated in the literature to date. Biomedical approaches predominate - they imply a direct relationship between disease and symptoms in the presence of illness. Other perspectives are less well-integrated within the literature; psychological approaches distinguish between 'bodily changes' (both sensations and visible changes) and 'symptoms' and are interested in the processes by which one becomes the other.[9,10,11] Sociological and anthropological approaches explore the processes by which bodily sensations are interpreted as symptoms; they are predicated on the view that symptoms evolve and are interpreted through the relationship between the individual and a given social and cultural context[12].

In the light of different purposes of the research, several theoretical models have been developed to describe the events and processes that underpin the pathway to symptomatic cancer diagnosis. Influential among these is the Danish model developed by Olesen and colleagues[4] and the Model of Pathways to Treatment (see Figure 1), a refinement by Walter and colleagues of the Andersen Model of Total Patient Delay.[13]

INSERT FIGURE 1 ABOUT HERE

Hence, early diagnosis research is characterised by its complexity, a poorly developed set of definitions and methodological tools, a lack of transparency in disciplinary perspectives and an absence of a widely-accepted underlying theoretical model. Accordingly, in order to advance the quality and consistency of studies of diagnostic intervals in symptomatic cancer, an international Consensus Working Group (CWG) was convened in November 2009. The purpose of the group was to formulate definitions of key time points and to make methodological recommendations for researchers in this field. The work was commissioned by Cancer Research UK and the

Department of Health in England, and undertaken under the auspices of the Cancer and Primary Care Research International (Ca-PRI) Network.[14]

The aim of this paper is to propose and discuss a standardised set of definitions that can be used in research on earlier cancer diagnosis, relating to key timepoints and intervals in early diagnosis research. We also present methodological recommendations, principally on data collection and analysis.

Methods

We used a 'consensus conference' approach[15] which encompasses consensus meetings, presentations and wide circulation of draft outputs for modifications and refinement. It has been used in other fields to produce guidance for researchers and clinicians, and is particularly useful when a wide range of clinical, sociological, cultural and other perspectives need to be incorporated within the guidance. We also drew on nominal group techniques[16] in our consensus meetings and in the exchange of information in the intervals between the meetings.

Members of the CWG were drawn from a range of disciplinary and methodological backgrounds including primary care, health services research, social science, psychology, epidemiology and statistics. The CWG was supported by an international expert reference group comprising individuals from a wide range of clinical and methodological backgrounds. The CWG met three times between January and November 2010 in London and Aarhus. Each of the meetings comprised a series of presentations addressing theoretical, empirical and research methodological key areas of the project; the ensuing discussions were recorded and circulated between meetings. Consensus was achieved through an iterative approach whereby proposed statements and checklists were developed and refined in the course of these meetings and informed by the presentations. Each element of the statement was discussed in detail by all members of the group until consensus was reached. The draft statement was then reviewed by an expert reference group and further refined in the light of their feedback.

The CWG had a close familiarity with the literature in this field. Nevertheless, we decided to underpin this by using systematic approaches to identify and assess relevant literature relating to existing questionnaires and survey tools. Accordingly, we further examined 279 included papers from a concurrent systematic review examining the association between outcomes and diagnostic intervals[17]. The inclusion criteria are shown in Box 1.

INSERT BOX 1 ABOUT HERE

Results

After examining the early cancer diagnosis literature, and taking into account the extensive theoretical work in this field, the CWG concluded that:

- there is little consistency in the definition and measurement of key time-points and intervals;
- there is little guidance for researchers in designing studies which require the measurement of diagnostic time-points and intervals;
- little work in this field explicitly uses a theoretical framework to underpin definitions and measurement of diagnostic intervals;
- there is a lack of transparency and precision over methods and instruments in early diagnosis research – typically, important study elements such as instrument development are poorly described.

Accordingly, the CWG group developed the 'Aarhus Statement' comprising the following series of definitions and recommendations:

1. Definitions

The relevance of individual time-points and intervals varies between health care systems. Nevertheless, in an international context, the importance of the following four time-points emerged in the Consensus Working Group discussions – we found the representation in Figure 2, illustrating the time-points and their associated intervals, helpful:

INSERT FIGURE 2 ABOUT HERE

Date of first symptom

This should be defined as *'the 'time-point when first bodily change(s) and/or symptom(s) are noticed'*. Researchers should consider this definition to encompass several key components: the *date the first bodily change was noticed, the date the first symptom was noticed, the date when the person perceives a reason to discuss the symptom with a health care professional, and the da*te *the first 'alarm' or 'higher risk' symptom was noticed.* Researchers should report clearly which of these definitions of first symptom has been applied in their study.

The term 'patient delay' should no longer be used - instead, 'appraisal interval' (time taken to interpret bodily changes/symptoms) and 'help-seeking interval' (time taken to act upon those interpretations and seek help) are more helpful in describing the 'patient interval' (see Figure 1). Researchers should acknowledge in their definition how they have dealt with the complexity of this time-point, for example, where symptoms are common, non-specific, multiple or chronic, where other morbidities co-exist, and the social and healthcare context. It is important to recognise that often symptoms are medically defined, and such definitions may be inconsistent with lay symptom definitions.

Date of first presentation

This should be considered 't*he time-point at which, given the presenting signs, symptoms, history and other risk factors, it would be at least possible for the clinician seeing the patient to have started investigation or referral for possible important pathology, including cancer'.* There should be a demonstrated understanding of the pattern of symptoms in the lead up to the first presentation (eg frequency, chronicity, presence of other symptoms), and whether the date has been defined from the perspective of the health care provider or the patient. There should be precise descriptions of where this first presentation occurs (eg primary care, hospital department etc).

Date of referral

This should be considered *'the time-point at which there is a transfer of responsibility from one health care provider to another (typically, in 'gatekeeper' health care systems, from a primary care provider to a doctor/service specialising in cancer diagnosis and management) for further clinical diagnostic and management activity relating to the patient's suspected cancer'.* Patients may be referred more than once or between specialists; there are risks of cross-referrals within secondary care, and complex diagnostic routes (where patients ultimately receive treatment from specialist services which are different to those initially targeted). Researchers should use a consistent and explicit method for dealing with such complexities. Further, referral for investigations should be considered as a subsidiary time-point that may be of significance in some health care settings, but is not equivalent in cases where actual responsibility for patient management is not transferred. The service targeted in the referral should also be described.

Date of diagnosis
Studies reporting any time interval that either begins or ends with 'diagnosis' should be explicit about how that date is measured and what it actually means with respect to the diagnostic journey within that health system. Researchers should consult the well-developed hierarchical rationales available in the public domain in choosing their definition of date of diagnosis – one example is the hierarchy produced by the European Network of Cancer Registries[18] which is shown in Box 2.

INSERT BOX 2 ABOUT HERE

---

2. Methodological Approaches

Most studies on early cancer diagnosis involve retrospective data collection, which can be subject to recall bias.[19] Prospective studies on patients with symptoms are often recommended[20], although they pose difficulties due to the large numbers of participants required. Time since diagnosis is a critical sampling issue - sampling too long after diagnosis increases the risk of recall bias and increases the likelihood of attrition due to death or terminal illness. Conversely, sampling too soon after diagnosis may be insensitive, and may increase the likelihood of attrition due to ongoing active treatment (surgery, chemotherapy, radiotherapy). The sampling method, whether through a cancer registry, hospital-based sampling or other mechanism, should be reproducible and thoroughly described.

Primary data collection from patients and/or providers
The key challenge in collecting data on diagnostic intervals is to obtain valid and complete data and, at the same time, capture the complexity of the pathways to a diagnosis of cancer. However, this might lead to overly-cumbersome approaches and significant response burden. Therefore, researchers should develop questions which are 1) precise about the time-point they are endeavouring to describe, 2) applicable to both the cancer in question and the likely symptom and symptom complexes that the patient experiences, and 3) specific about the context of the health care system.

Questions should derive from a clearly-stated theoretical basis, and the health care context should be clearly described, to allow for appropriate interpretation of responses. Ways in which the measurement approach takes into account multiple symptoms, chronic symptoms and co-existent co-morbidity should be described.

For measurement of the date of first symptom and date of first presentation, in-depth qualitative interviews with patients are preferred when there is a need for detailed understanding of the time-point. Strategies such as calendar-landmarking[21] should be considered to reduce recall bias in qualitative interviews. In common with research in other fields, patient-completed surveys (typically in the form of self-completion questionnaires) provide the best opportunity for producing large, population-based data-sets, but their limitations in capturing the complexity of these time-points should be acknowledged and researchers should clarify how the survey has been developed in accordance with international standards.[22]

For date of first presentation, ideally information should be gathered from both patients and primary care providers, as their concepts of this time-point may differ, particularly in the context of vague, non-specific or chronic symptoms. Open-ended questions, while making coding more difficult, are typically needed to encompass the complexity of cancer-related symptoms when asking patients.

It is important to scrutinise the literature to be sure that projects include the best available measurement instrument. In future we might see internationally validated standard questionnaires in this field. Therefore, it is also important that researchers provide very transparent information on development of interviews and surveys.

Case-note audit
Retrospective examination of case-notes (eg charts, medical records) is an important source of information in medical research and clinical audit; a vital consideration is the accuracy of this information.[23] This approach should ideally be used to augment or validate data from other sources – particularly for complex time-points such as date of first symptom. There should be a description of the process of how the clinician makes and codes the record, and how clinicians interpret and record the patient's history in the clinical notes. Further, the likely completeness of the case-note data should be described. There is typically little data in secondary care notes on early stages of diagnostic pathways (ie time-points before the patient arrives in secondary care), although they are an important source of information on latter stages of the patient journey.

Primary care database analysis
There has been a growth in the number and quality of databases which collect encounter and diagnostic information from primary care – one example is the General Practice Research Database (GPRD).[24] These offer the potential to conduct analyses on large samples, and can facilitate study designs such as retrospective cohort studies. When used, there should be a thorough description of the database and its capacity to capture valid information on time-points and intervals (eg completeness and accuracy of recording of encounter data). Further, the limitations deriving from coding systems used in the database should be described and there should be standard procedures for both conceptualising data fields and uploading data in a standard format. In common with case-note audit, there are limitations in deriving information on time-points such as date of first symptom from analysis of databases – although both audit and database analysis can be useful in deriving information on date of first presentation.

3. The Aarhus Check-list
Based on these recommendations, the CWG has developed a check-list for early diagnosis researchers. The check-list is intended for individuals undertaking research which involves the description and measurement of intervals in the cancer diagnostic journey. It has been produced with the intention of promoting greater consistency and transparency in methods and measurement. Furthermore, it is a resource both for those developing studies which require measurement of intervals and/or mapping of cancer patient journeys, and for journal paper reviewers, editors and funding bodies who can use these items as a framework for assessing the quality of research applications and papers.

INSERT AARHUS CHECK-LIST ABOUT HERE

Discussion
We report the first comprehensive guidance for the conduct and reporting of research in this field. We have highlighted the many needs and pitfalls in early diagnosis research, and call for greater precision and transparency in both definitions and methods. Our rigorous and multidisciplinary approach to consensus development, coupled with detailed analysis of the existing early diagnosis literature, has produced guidance which can promote [the primary] consistency and methodological rigour in this field. Our recommendations are relevant to researchers, funding bodies, and journal editors and reviewers. The next step is to examine whether these definitions and recommendations can be readily adopted by researchers; evidence from similar previous initiatives suggests that researchers are receptive to new methodological guidance, and that it can have a positive effect on research quality[25, 26].

While our check-list has not addressed analytical techniques, the choice of technique is crucial given that analyses of time intervals are used to monitor quality of clinical trajectories and in

research on prognosis of cancer. In prognostic research there is strong evidence to suggest that time interval data should be analysed in a statistical model using time interval as a continuous variable, rather than a dichotomized model, in order to minimise bias arising from what has been termed 'the waiting time paradox'.[27,28]

We have identified a pressing need for more methodological work in this field and we are in the process of further evaluation and validation of our recommendations, which can be viewed in greater detail on the Ca-PRI website.[14] Feedback on the guidance will be used for future updates.

Conclusions

There are growing international efforts to describe and measure patient journeys prior to a cancer diagnosis. Accurate descriptions of these patient journeys and valid measurements of diagnostic intervals are essential to determine the effectiveness of interventions to reduce them. The Aarhus Check-list will facilitate standardised and uniform definition and reporting of studies in this area.

Box 1: Inclusion criteria for systematic review

A study or analysis was included in the review if it fulfilled all the criteria below:
- Participants: only studies or analysis that examined symptomatic cancer patients presenting with primary cancers were included.
- Aim of study: ideally, to analyse the impact of the interval between the time of first symptomatic presentation to definitive diagnosis or treatment on a health-related outcome.
- Outcome: any study analysing any health-related outcome was considered for inclusion; these include survival, diagnostic stage, treatment assignment after diagnosis, and other measures of "curability". Survival was the main outcome of interest; additionally, studies examining health-related quality-of-life, resource-use and psychological outcomes were considered for inclusion.
- Study design: there were no restrictions on the types of study design that were considered for inclusion.

Box 2 - European Network of Cancer Registries: Hierarchy for defining date of diagnosis[18]

In order of declining priority…..

1. Date of first histological or cytological confirmation of this malignancy (with the exception of histology or cytology at autopsy). This date should be, in the following order:

   a) date when the specimen was taken (biopsy)

   b) date of receipt by the pathologist

   c) date of the pathology report.

2. Date of admission to the hospital because of this malignancy.

3. When evaluated at an outpatient clinic only: date of first consultation at the outpatient clinic because of this malignancy.

4. Date of diagnosis, other than 1, 2 or 3.

5. Date of death, if no information is available other than the fact that the patient has died because of a malignancy.

6. Date of death, if the malignancy is discovered at autopsy.
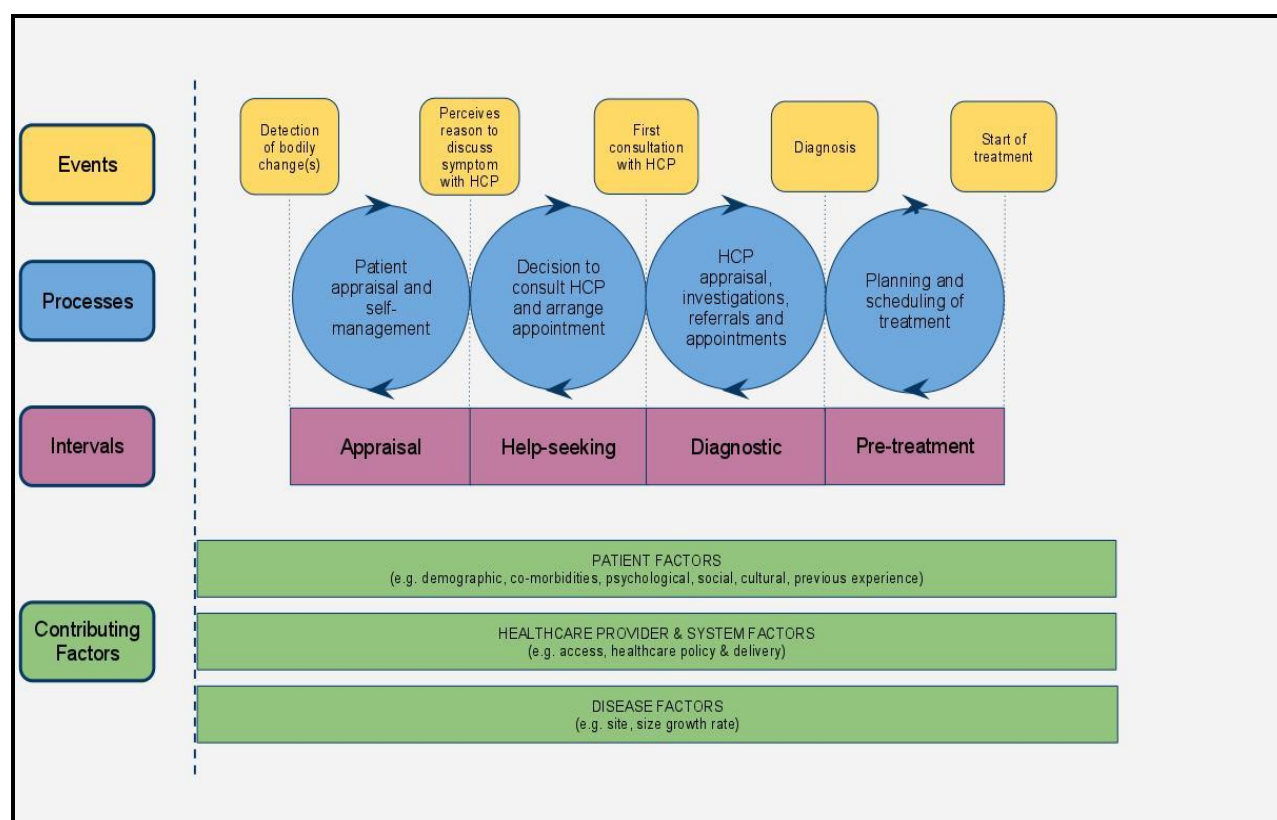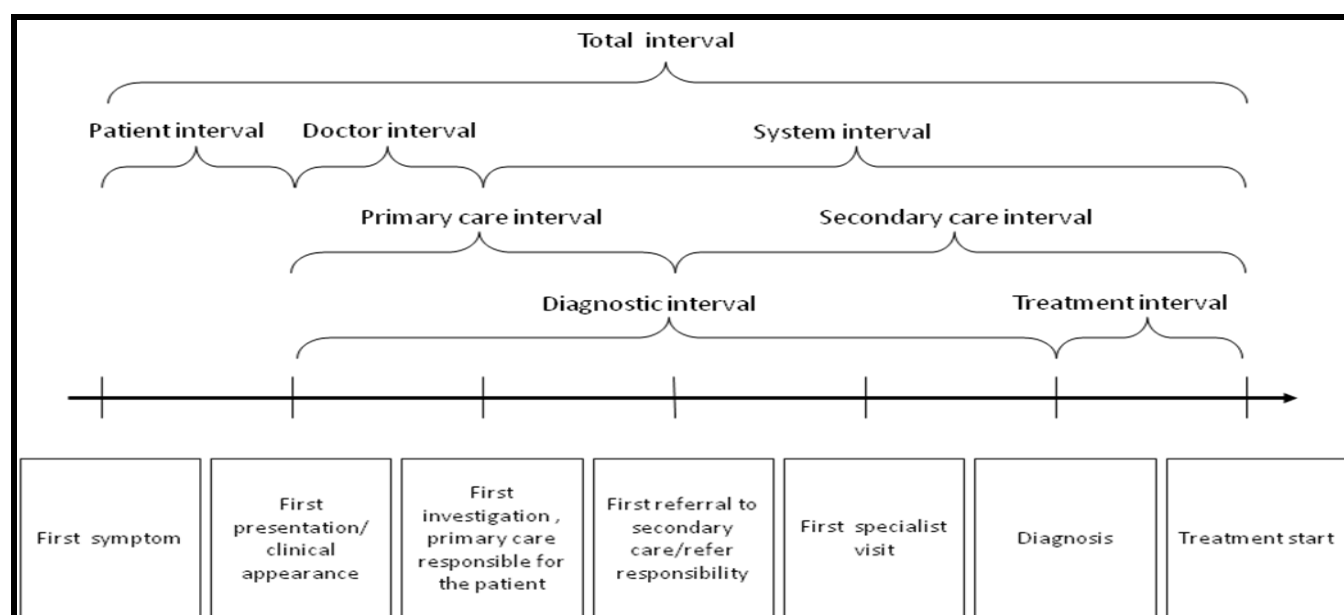
Figure 1 - Model of Pathways to Treatment13

Figure 2 - An illustration of the overall milestones and time intervals in the route from first symptom until start of treatment[4]

The Aarhus Check-list

Item		Y/N
DEFINITIONS OF TIME-POINTS AND INTERVALS		
1.	For studies requiring the measurement of an interval, are the beginning and end points of this interval clearly defined?	
2.	For all time-points and intervals described are there precise, transparent and repeatable definitions, and is the complexity of time-points such as date of first symptom and date of first presentation addressed?	
For studies which require an estimate of date of first symptom:		
3.	Do the researchers refer to a theoretical framework underpinning definition of this time-point?	
4.	Is there a discussion of the different biases influencing measurement of this time-point	
For studies which require measurement of a date of first presentation to healthcare:		
5.	Do the researchers discuss the complexity of the date of first presentation?	
For studies which require measurement of a date of referral:		
6.	Do the researchers discuss the nature of the referral, and provide adequate detail – eg whether it was for investigation or consultation by a colleague in secondary care?	
For studies which require measurement of date of diagnosis:		
7.	Do the researchers use an existing hierarchical rationale for date of diagnosis measurement?	
MEASUREMENT		
8.	Is the healthcare context in which the study is based fully described?	
9.	Do the questions on time-points and/or intervals clearly derive from stated definitions?	
10.	Do researchers acknowledge the need for theoretical validation and make reference to the theoretical framework(s) underpinning measurement and analysis of the time-points?	
For studies using questionnaires and/or interviews with patients and/or healthcare providers:		

11.	Has a validated instrument been used?	
12.	Have the researchers included a copy of their instrument?	
13.	Is there some discussion of how reliability and validity (trustworthiness) has been established?	
14.	Do researchers acknowledge the need for theoretical validation and make reference to the theoretical framework(s) underpinning measurement and analysis of the time-points?	
15.	Is there discussion of the different biases influencing measurement of the time-points, such as how and when the question is asked and who is being asked?	
16.	Is the timing of the interview in relation to the date of diagnosis provided?	
17.	Is there any triangulation of self-reported data with other data sources such as case notes?	
18.	Is data analysis described in full including how and why data are categorised, how missing, and incomplete data are managed and how outliers at both ends of the spectrum are accounted for?	
For studies using primary case note audit and database analysis:		
19.	Case-note analysis: is there a clear and precise description of how case note data were used to ascertain time-points with an acknowledgment of limitations of such data?	
20.	For database analysis: is there a thorough description of the database chosen including sampling coverage and completeness of information?	

References

1. Richards, M.A. The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *Br J Cancer* **101 Suppl 2**, S1-4 (2009).
2. Rubin, G., Vedsted, P. & Emery, J. Improving cancer outcomes: better access to diagnostics in primary care could be critical. *Br J Gen Pract* **61**, 317-318 (2011).
3. Coleman, M.P.*, et al.* Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* **377**, 127-138 (2011).
4. Olesen, F., Hansen, R.P. & Vedsted, P. Delay in diagnosis: the experience in Denmark. *Br J Cancer* **101 Suppl 2**, S5-8 (2009).
5. Cancer Reform Strategy. *Department of Health* (2007).
6. Jordan, S.J.*, et al.* Pathways to the diagnosis of epithelial ovarian cancer in Australia. *Med J Aust* **193**, 326-330 (2010).
7. Fisher, D.A.*, et al.* Determinants of medical system delay in the diagnosis of colorectal cancer within the Veteran Affairs Health System. *Dig Dis Sci* **55**, 1434-1441 (2010).
8. Neal, R.D. Do diagnostic delays in cancer matter? *Br J Cancer* **101 Suppl 2**, S9-S12 (2009).
9. van Wijk, C.M. & Kolk, A.M. Sex differences in physical symptoms: the contribution of symptom perception theory. *Soc Sci Med* **45**, 231-246 (1997).
10. Bradley, E.J., Calvert, E., Pitts, M.K. & Redman, C.W.E. Illness Identity and the Self-regulatory Model in Recovery from Early Stage Gynaecological Cancer. *Journal of Health Psychology* **6**, 511-521 (2001).
11. Teel, C.S., Meek, P., McNamara, A.M. & Watson, L. Perspectives unifying symptom interpretation. *Image J Nurs Sch* **29**, 175-181 (1997).
12. Alonzo, A.A. An illness behavior paradigm: a conceptual exploration of a situational-adaptation perspective. *Soc Sci Med* **19**, 499-510 (1984).
13. Walter, F., Webster, A., Scott, S. & Emery, J. The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. *J Health Serv Res Policy* **In press**(2011).
14. http://www.ca-pri.com/.
15. Durocher, A. [Consensus conference: methodology and contribution]. *Therapie* **49**, 185-190 (1994).
16. Black, N.*, et al.* Consensus development methods: a review of best practice in creating clinical guidelines. *J Health Serv Res Policy* **4**, 236-248 (1999).
17. Neal, R. Time to diagnosis in symptomatic cancer: Does it have an effect on clinical outcomes? Systematic review. *National Awareness of Cancer and early Diagnosis (NAEDI)* (London, 2011) http://info.cancerresearchuk.org/spotcancerearly/naedi/AboutNAEDI/Researchevaluationandmonitoring/first-naedi-research-conference/.
18. http://www.encr.com.fr/incideng.pdf. European Network of Cancer Registries (ENCR)
19. Friedman, G.D., Skilling, J.S., Udaltsova, N.V. & Smith, L.H. Early symptoms of ovarian cancer: a case-control study without recall bias. *Fam Pract* **22**, 548-553 (2005).
20. Summerton, N. Symptoms of possible oncological significance: separating the wheat from the chaff. *Bmj* **325**, 1254-1255 (2002).
21. Glasner, T. & van der Vaart, W. Applications of calendar instruments in social surveys: a review. *Quality and Quantity* **43**, 333-349 (2009).
22. Mokkink, L.B.*, et al.* The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* **19**, 539-549 (2010).
23. Lilford, R.*, et al.* Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Policy* **12**, 173-180 (2007).
24. Jones, R., Charlton, J., Latinovic, R. & Gulliford, M.C. Alarm symptoms and identification of non-cancer diagnoses in primary care: cohort study. *Bmj* **339**(2009).

25. Wynne, K.E.*, et al.* Results of a longitudinal study of rigorous manuscript submission guidelines designed to improve the quality of clinical research reporting in a peer-reviewed surgical journal. *Journal of Pediatric Surgery* **46**, 131-137 (2011).
26. Willis, B. & Quigley, M. The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Med Res Methodol* **11**, 163 (2011).
27. Crawford, S.C.*, et al.* The waiting time paradox: population based retrospective study of treatment delay and survival of women with endometrial cancer in Scotland. *Bmj* **325**, 196 (2002).
28. Torring, M.L.*, et al.* Time to diagnosis and mortality in colorectal cancer: a cohort study in primary care. *Br J Cancer* **104**, 934-940 (2011).