Malaysian J Pathol 2018; 40(3) : 319 – 323

ORIGINAL ARTICLE

Heterogeneity in studies of p16 in cervical lesions in different Malaysian institutions: Time to consider collaborative study

Manimalar Selvi NAICKER[1], Geok Chin TAN[2]

[1]Department of Pathology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia and [2]Department of Pathology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

Abstract

Introduction: Clinical decision making becomes difficult when clinical and methodological heterogeneity does not permit synthesis of results of multiple small studies. *Aim:* For studies done in Malaysia, to identity the sample sizes and heterogeneity present in the various studies which used p16 in evaluating lesions of the cervix. To evaluate if it would be possible for a single study to answer the various questions posed by the original authors. To highlight areas where the design features of future studies can be optimised. *Materials and Methods:* Various databases were searched using synonyms for p16 AND cervix AND Malaysia. These were assessed for broad conformity to a Diagnostic Test Accuracy format. Methodological and clinical heterogeneity indicators were extracted into standardised fields. *Results:* There were 5 studies eligible for inclusion. Each sought to study different aspects of the disease such as diagnostic grade stratification and pathogenesis. The study type broadly conformed to a Diagnostic Test Accuracy format. The study design used was either consecutive or non-consecutive. Sample size ranged from 75 to 201. Clinical heterogeneity was present in the choice of controls with some using normal and some using inflamed tissue. Methodological heterogeneity in applying the reference test, index test and different antibody clones were present. *Conclusion:* There was both clinical and methodological heterogeneity making synthesis of studies difficult. It is possible to design a study which would be able to answer all the questions posed by the original authors with internal validity while at the same time increasing sample size.

Keywords: p16, cervix, heterogeneity

INTRODUCTION

Medical doctors practising in Malaysia are increasingly faced with the task of making sense of a larger number of methodologically (protocol) and clinically (population) heterogeneous studies when making clinical decisions for our population of patients. It is time consuming to do systematic reviews for the many clinical questions that we face on a daily basis. Further, it ignores the fact that systematic reviews are often clinically not useful as the primary studies included may have clinical and methodological heterogeneity. Malaysia has a system of hospitals comprised the Ministry of Health (MOH), the Ministry of Higher Education (MOHE) and Private Hospitals. This has often resulted in single-centre, small sample size and clinically and methodologically heterogeneous studies. It is possible to avoid this problem by forming national consortiums and undertaking jointly-designed and larger multi-centre studies. However, often, different researchers seek to answer slightly different questions making a national consortium a potentially contentious issue.

In histopathology, the Diagnostic Test Accuracy (DTA) type study is a major workhorse. Taking advantage of archival tissue available, it is used to answer questions such as sensitivity and specificity of an immunohistochemical marker (clinical), testing and standardisation of new biomarkers (translational) and answer pathogenesis type questions (pre-clinical). Hence, it is often possible to unify the study design for many types of research questions so long as the biomarkers and tissue being

Address for correspondence: Dr Manimalar Selvi Naicker, Department of Pathology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. Tel: 03-79675797. Fax: 03-79556845. Email: mala@ummc.edu.my

interrogated are the same and the study design broadly conforms to that used in DTA studies.

The aims of this study were: 1) to assess the sample sizes and clinical and methodological heterogeneity in studies using the p16 antibody on cervical lesions; 2) to assess if a single study can be designed to answer questions posed by various researchers; 3) to identify areas where future studies can be strengthened in terms of study design.

MATERIALS AND METHODS

Definition

For the purpose of this study, clinical heterogeneity refers to variability in the participants (types of cases) studied. Methodological heterogeneity refers to variability in study design (protocol). Clinical and methodological heterogeneity may lead to statistical heterogeneity which in turn is defined as the variability in the intervention effects being evaluated in the different studies.[1]

Search strategy

PubMed, Scopus and Google Scholar were searched systematically using very broad keyword searches for a sample of Malaysian studies. *((Malaysia) AND cervi*) AND ((((p16) OR p16 INK4a) OR p16 INK4) OR CDKN2A)*. In order to avoid known limitations of Boolean and MeSH terms searches, individual combinations were also searched using (p16, $p16^{INK4a}$, $p16^{INK4}$, CDKN2A) AND cervi* AND Malaysia. A total of 8 studies yielded 5 eligible studies published between 2010 and 2014. The searches were made during the period 01/01/2016 to 31/01/2016. The search was repeated again in September 2018. No new unique studies using tissue sections were found during the interim period.

Inclusion criteria: Malaysian patients, studies using the p16 immunohistochemical stain, cervical tissue being evaluated, using tissue sections, broadly conform to Diagnostic Test Accuracy type studies. Exclusion criteria: Cytology specimens.

Data was extracted into pre-defined categories which reflected areas of possible clinical and methodological heterogeneity. Concordance between stated aim and final conclusion will be reviewed to evaluate if it was the authors intention that the studies should broadly fit into the Diagnostic Test Accuracy framework.

RESULTS

There was a total of 5 papers which fit the inclusion and exclusion criteria and the extracted data in Table 1. The studies in order of publication were Tan *et al*. (2010)[2], Cheah *et al*. (2011)[3], Looi *et al*. (2011)[5] and Krishnappa *et al*. (2014)[6]. All studies were university based with some collaborating with Ministry of Health hospitals. Tan *et al*.[2] and Looi *et al*.[5] studies were from the same university. Of this, Cheah *et al*. (2012)[4] was an extension of a previous study (Cheah *et al*. 2011)[3] whereby another 23 cases were added to the squamous cell carcinoma group. The methodology and period of case selection for both was similar, hence for the purpose of methodology the study will be referred to as Cheah *et al*. (2011)[3] but the data will utilise the larger Cheah *et al*. (2012)[4] study to avoid duplication and double counting.

Comparison of aims and conclusions of the studies

In Tan *et al*. (2010)[2] paper, the aim was to determine the expression of p16INK4A as possible diagnostic biomarker in cervical squamous neoplasm. The author concluded that p16 INK4A expressions correlated well with the increasing grade of CIN. In the Cheah *et al*. (2011)[3] paper, the author aimed to use p16 expression as a surrogate marker of E2F release and G_1S transition. Thus, they expected to see higher expression in the SCC and HSIL groups. They also attempted to further understand the role of E2F (using surrogate p16), in the evolution of HSIL to SCC. They concluded that the study showed that p16 expression was much higher in HSIL and SCC- and fits with the theory. Looi *et al*. (2011)[5] paper aimed to identify biomarkers of cellular proliferation in cervical lesions by detecting the expression of p16. In conclusion, they observed an increase in expression of p16 during the progression of cervical cancer. Krishnappa *et al*. (2014)[6] paper aimed to evaluate the diagnostic value of p16 IHC in diagnostic of the different categories of HR-HPV infected cervical lesions. The study concluded that p16 was useful in distinguishing between high- and low-grade lesions.

As all the studies intended to grade the percentage of cases in each category which were p16 positive, they were included in the review. All the studies have a similar aim, to determine the usefulness of p16 expression in assessing cervical neoplasms. However, each study differs slightly in that some intended to use

Table 1: Data extracted from the selected studies: General features and results

Year published	First author	Hospitals	Archival tissue was likely used	Year of sample recruitment	Antibody clone/ Manufacturer	Results (% of p16 positivity)	Sample size	Other stains
2010	Tan *et al*.[2]	UKM, HKL	Yes	Jan 2003 to Dec 2007	NA/ NeoMarkers	Normal (0), CIN1 (25.4), CIN2 (42.9), CIN3 (95.9), SCC (98.6)	201	Survivin
2011	Cheah *et al*.[3]	UM	Yes	Jan 2006 to Dec 2008	NA/CINtec Histology Kit	Adjacent Normal (0), CIN1/LGSIL (3.4),	109	No
2012	Cheah *et al*.[4]	UM	Yes	Jan 2006 to Dec 2008		CIN2/3/HGSIL (88.9), SCC (86.8)		
2011	Looi *et al*.[5]	UKM	Yes	NA	E6H4, Dako, Denmark	Normal (0), CIN1/LGSIL (25), CIN2/3/HGSIL (50), SCC (62.5)	131	Ki67
2014	Krishnappa *et al.*[6]	Hosp. Seremban/ IMU	Yes	Jan 2012 to Dec 2012	2D9A12/ Abcam ab 54210	Chronic cervicitis (0), CIN (72), SCC (100)	75	HPV16Li

NA – Not available, UKM – Universiti Kebangsaan Malaysia, UM – Universiti Malaya, HKL – Hospital Kuala Lumpur, IMU – International Medical University, CIN – Cervical intraepithelial neoplasia, SCC – Squamous cell carcinoma

p16 as a diagnostic marker to stratify between grades, while some determine to understand the underlying pathophysiology of progression of cervical cancer.

Assessment for heterogeneity

All studies used tissue and not cytology smears.

Patient population and target condition: Cervical lesions (normal, inflamed, *in situ* neoplasia and invasive neoplasia) in Malaysian patients.

Reference Test: The CIN system usually used in histological tissue sections is based on cell as well as architectural changes whereas the SIL system is used for cytological smears and uses mainly cell changes. The SIL system is used as a screening tool whereas the CIN system is used as definitive assessment. There is a correspondence between CIN and SIL systems, LGSIL (CIN 1) and HGSIL (CIN 2/3), but they are not considered identical.[7,8]

Tan *et al*. (2010)[2] and Krishnappa *et al*. (2014)[6] studies used the conventional terminology (CIN/ SCC), whereas the Cheah *et al*. (2011)[3] and Looi *et al*. (2011)[5] studies used CIN/SIL terminology interchangeably. All authors did not explicitly state the source of the diagnostic criteria used to assign cervical tissue to the various categories on H&E. There was no further follow up to determine if the seemingly normal/inflamed cases exhibited dysplasia on subsequent cytology or the CIN cases presented with metastasis at follow up indicating a possible misclassification bias.

The possible reason for the difference in terminology used by different groups may be the change in terminology during that period, from a three-tiered grading system cervical intraepithelial neoplasia (CIN) 1, 2 and 3 to a two-tiered grading low-grade and high grade squamous intraepithelial lesion (LSIL and HSIL).[9] Tan *et al*. (2010)[2], Cheah *et al*. (2011)[3] and Looi *et al*. (2011)[5] assessed neoplastic lesional tissue against normal tissue. In contrast, Krishnappa *et al*. (2014)[6] assessed neoplastic lesional tissue against inflamed tissue.

Index Test: The studies used monoclonal antibodies from different manufacturers. Actual staining protocols were not further compared. The Cheah *et al*. (2011)[3] study required either nuclear or cytoplasmic staining whereas the

others required both for staining pattern. Krishnappa *et al*. (2014)[6] study quantified staining intensity using the Allred system. The others did not mention the staining intensity. The Tan *et al*. (2010)[2] and Looi *et al*. (2011)[5] studies classified negative staining as those with less than 5% of cells staining positive. Cheah *et al*. (2011)[3] considered less than 75% as negative and greater than 75% as positive. It is not clear how the Krishnappa *et al*. (2014)[6] study assigned positive and negative categories as online searches indicated that there could be more than one way to dichotomise this score.

Study Design: Though not explicitly stated in all studies, archival material seems to have been used. Diagnostic Test Accuracy type studies that have used either a consecutive or non-consecutive design with control being non-neoplastic tissue (normal or inflamed tissue) and cases being the various subgroups of *in situ* (CIN) and invasive (SCC) neoplasia cases. The sample size ranged from 75 to 201 with a total of 123 normal or inflamed tissue, 250 CIN and 198 SCC. Overall, tissues from 516 individual patients was used in this study (though the effective sample size was 571).

Combined p16 expression results in different categories:
Normal and inflamed tissue – Negative
CIN 1 – 3.4 to 25% positivity
CIN 2/3 – 50 to 89% positivity
SCC – 63 to 100% positivity

In combined CIN/SIL as one group the positivity ranged from 40 to 72%. The percentage of positivity was higher in SCC than CIN in all studies, except study by Cheah *et al*. (2011)[3].

DISCUSSION

The patient population and target condition are appropriate for this study (as these were determined at the outset as inclusion criteria). However, overall the studies show clinical heterogeneity in the choice of controls with some using normal and some using inflamed tissue. Methodological heterogeneity in applying the reference test, index test and different antibody clones were noted. Since both clinical and methodological heterogeneity were present, the results of a meta-analysis would not be meaningful. The broad observation is that in all studies normal/inflamed tissue did not stain positive and hence p16 staining could serve as a useful marker in diagnostically difficult cases in distinguishing neoplastic from non-neoplastic

Comment
There are several benefits if similar studies are conducted as a single multi-centre study.

(1) The study design could be either consecutive or non-consecutive.
(2) A consecutive design, in addition to Specificity and Sensitivity, will also allow a Positive and Negative Predictive Value to be calculated since it incorporates Prevalence.
(3) The Reference Test can be standardised and source of diagnostic criteria cited and thereafter used nationally.
(4) Inter and intra observer reliabilities can be assessed for the H&E stains.
(5) For each proposed grading/quantification system, reliability using inter and intra observer reliability studies can be evaluated.
(6) Both the reference test and index test can be interpreted without knowledge of the other (blinding).
(7) A systematic review can be done to evaluate the various antibody manufacturers prior to selecting one or two for comparison.
(8) The consistency of technical performance of the immunohistochemical stains across various laboratories can be evaluated.
(9) All the above would be in addition to the percentage in each category which are positive or negative as done in the original studies.

It may answer the pathogenesis type question posed by Cheah *et al*. and site-specific studies could have been planned using the other stains Looi *et al*. (Ki-67)[5], Tan *et al*. (survivin)[2] and Krishnappa *et al*. (HPV 16 L1)[6] and used to answer other questions. In conclusion, there was both clinical and methodological heterogeneity making synthesis of studies difficult. However, small studies may not be avoidable as each institution only has a small sample size. Standardisation of study protocol and study population will make future systematic reviews and meta-analysis possible.

ACKNOWLEDGEMENT

General research facilities provided by University of Malaya used for this study.

Conflicts of Interest: The authors declared no conflict of interest.

REFERENCES

1. The Cochrane Collaboration. What is heterogeneity? 2011; 9.5.1 [updated 201113/09/2017]. Version 5.1.0 [http://handbook-5-1.cochrane.org/] (Accessed on 15/11/2018)
2. Tan GC, Norlatiffah S, Sharifah NA, Razmin G, Shiran MS, Hatta AZ. Immunohistochemical study of p16 INK4A and survivin expressions in cervical squamous neoplasm. Indian J Pathol Microbiol. 2010; 53(1): 1-6.
3. Cheah PL, Looi LM, Mun KS, Abdoul Rahman N, Teoh KH. Implications of continued upregulation of p16(INK4a) through the evolution from high-grade squamous intraepithelial lesion to invasive squamous carcinoma of the cervix. Malays J Pathol. 2011; 33(2): 83-7.
4. Cheah PL, Looi LM, Teoh KH, Mun KS, Nazarina AR. p16(INK4a) is a useful marker of human papillomavirus integration allowing risk stratification for cervical malignancies. Asian Pac J Cancer Prev. 2012; 13(2): 469-72.
5. Looi ML, Ahmad Zailani Hatta MD, Siti Aishah MA, Yasmin AMY. Expression of p16 and pKi-67 in Cervical Preneoplasia and Neoplasia. Asia-Pacific Journal of Molecular Medicine. 2011; 1: 1-6.
6. Krishnappa P, Mohamad IB, Lin YJ, Barua A. Expression of P16 in high-risk human papillomavirus related lesions of the uterine cervix in a government hospital, Malaysia. Diagn Pathol. 2014; 9: 202.
7. Kumar V, Abbas A, Aster J. Robbins & Cotran Pathologic Basis of Disease. 10th edition, USA. Saunders; 2017.
8. Abnormal Cervical Cancer Screening Test Results 2016 [http://www.acog.org/Patients/FAQs/Abnormal-Cervical-Cancer-Screening-Test-Results] (Accessed on 15/11/2018)
9. Waxman AG, Chelmow D, Darragh TM, Lawson H, Moscicki AB. Revised terminology for cervical histopathology and its implications for management of high-grade squamous intraepithelial lesions of the cervix. Obstet Gynecol. 2012; 120: 1465-71.