Abstract

Introduction: Cervical cancer screening programs all over the world are being radically revised; preventive screening is increasingly based on detecting infection with human papillomavirus (HPV). This change was driven by large international studies (comparing cytological and HPV-based screening), which were then summarized in systematic reviews or HTA reports. We carried out a systematic comparison of these meta-analyses, comparing their contents, quality, results and recommendations.

Material and Methods: A systematic search in MEDLINE identified eight meta-analyses. Seven more papers were also included in our analysis. The quality of these publications was reviewed using the AMSTAR criteria and presented in tabular form.

Results: There were significant differences with regard to the primary studies included in the meta-analyses, the quality of the meta-analyses, the endpoints, and the outcomes and recommendations based on these endpoints.

Conclusion: Different meta-analyses on the same or similar issues sometimes came to quite different conclusions because they used different approaches. This is quite significant because the underlying primary data did not change. The data indicated that HPV-based screening was superior. The heterogeneity of the meta-analyses affects the decisions taken by policymakers in the healthcare system.

Zusammenfassung


Material und Methoden: Durch eine systematische Suche in MEDLINE konnten 8 Metaanalysen identifiziert werden. Sieben weitere Arbeiten wurden in die Analyse eingeschlossen. Die Qualität wurde mittels AMSTAR überprüft und tabellarisch dargestellt.

Ergebnisse: Es zeigten sich teilweise deutliche Unterschiede im Hinblick auf die eingeschlossenen Primärstudien, die Qualität der Metaanalysen, die Endpunkte und auch die Ergebnisse und die daraus resultierenden Empfehlungen.

Introduction

Opportunistic cervical cancer screening programs based on annual Pap smears were rolled out in Germany in 1971 and led to a significant decrease in the incidence of cervical cancer by 60–70% [1]. Despite this success, this approach has certain weaknesses. For example, it is currently not possible to comprehensively document and evaluate participation rates, although this significantly affects the future incidence of invasive carcinoma [2]. Other disadvantages of the current screening program for cervical cancer include the lack of a uniform workup procedure and the absence of continuous quality controls.

In addition to the proposal to introduce an organized screening program, the current discussion on the need to revise cervical cancer prevention demanded by politicians (National Cancer Plan and the Law on the Early Detection and Registration of Cancer [Krebsfrüherkennungs- und -registergesetz]) has also focused on the question whether screening should be primarily based on testing for HPV. The new S3 guideline “Prevention of Cervical Cancer” has also focused on this issue (http://leitlinienprogramm-ontologie.de/Zervixkarzinom-Praevention.89.0.html). These considerations were prompted by several large randomized controlled studies carried out in recent years which appeared to show that screening for HPV was superior to early detection based on cytology (summarized in [3]). In many countries, these studies served as the basis for a revision of cervical cancer screening programs. In a number of countries, healthcare policymakers requested that systematic reviews or meta-analyses be carried out to compare HPV with cytology for the prevention of cervical cancer and subsequently based their healthcare policy recommendations on the outcomes of these analyses.

Interestingly, these meta-analyses often differed significantly from one another; the different analyses included or excluded different primary studies and evaluated and assessed effects and overall results differently. This study aimed to review the methodological differences between the various meta-analyses in a systematic reappraisal of the currently available aggregated evidence on the issue of HPV vs. cytology and to compare the different results of these analyses in order to show the basis for the different recommendations issued in different countries.

Material and Methods

Search strategy

A comprehensive systematic search* for systematic reviews of evidence was carried out in MEDLINE via PubMed (http://www.ncbi.nlm.nih.gov/pubmed). The search criteria were, first and foremost, to find systematic reviews which compared cervical cancer screening using HPV test with cytological screening. This study also looked at systematic reviews which investigated individual aspects of HPV-based screening for cervical cancer. Non-systematic reviews, guidelines, best practice recommendations or primary studies were not included.

Analysis of the quality of the study

After summarizing the contents (the primary studies included, the method, and the results) the methodological quality of the systematic reviews were evaluated and compared using the so-called AMSTAR criteria (http://amstar.ca/Amstar_Checklist.php) [4].

Results

Literature search

After reviewing the abstracts, nine publications (eight independent studies; one of the studies was published in both a long and a short version) were identified out of the original 216 hits which complied with the search criteria [5–13]. Seven other systematic reviews which were not identified in the search – inter alia, two papers by the Institute for Quality and Efficiency in Healthcare (IQWIG) on the “Benefit assessment of HPV test in primary screening for cervical cancer” published in 2011 and 2014 (Table 1) as well as the analysis by Jos Kleijnen compiled for the S3 guideline on the prevention of cervical cancer – were also included in our review ([14–20]; a detailed summary of their contents is available in the online supplement. Only 13 of the included 15 meta-analyses focused on comparing HPV test with cytology in primary screening. Peirson et al. compared cytology, liquid-based cytology and HPV test with no screening [10], and Luu et al. evaluated the sensitivity and specificity of two different HPV detection methods (Hybrid Capture 2 and PCR) [6].

Study quality

The quality of the studies was compared using the AMSTAR criteria and is summarized in Table 2. The difference in the quality of the studies was sometimes quite significant, although it is important to point out that many methodological aspects may often not find mention in the final publication and the study was therefore evaluated as though these methodological aspects did not exist, even though it is possible that they were taken into consideration. The highest quality studies were the two studies by the IQWIG, along with a few others. There were also considerable differences in the choice of primary endpoints:

- histologically verified CIN 3+ [5]
- relative rate of detection, relative specificity, relative PPV [11]
- cervical cancer mortality, incidence of cervical cancer [10]
- CIN 2+, CIN 3+ [9]
- relative rate of CIN 2, CIN 2+, CIN 3+; cervical cancer mortality, and incidence of cervical cancer [7]
- overall survival, disease-specific (tumor-specific) survival, incidence of invasive cervical cancer, incidence of CIN 3/CIS, incidence of CIN 3+, injury arising directly or indirectly from screening, health-related quality of life, psychosocial aspects [15]

Table 1 Formulation of the benefit analysis by the IQWIG.

<table>
<thead>
<tr>
<th>IQWIG assessment</th>
<th>German guidelines</th>
<th>G-BA level of recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence of benefit</td>
<td>LoE Ia</td>
<td>Strongly recommended</td>
</tr>
<tr>
<td>Indications</td>
<td>LoE Ia</td>
<td>Recommended</td>
</tr>
<tr>
<td>Signs</td>
<td>LoE Ia–b</td>
<td>Qualified recommendation</td>
</tr>
<tr>
<td>No signs</td>
<td>&lt; LoE I</td>
<td>Not recommended</td>
</tr>
</tbody>
</table>

* Search strategy: PubMed, 24.09.2015. Search string: (cytology OR pap) AND (hpv OR papillomavirus) AND (“cervix neoplasm” or “cervical intraepithelial neoplasia” or “cervix dysplasia” or “cin” or “cervical cancer”). Limits: meta-analysis, systematic reviews. Number of hits: 216
Table 2  Evaluation of systematic reviews using the AMSTAR criteria (http://amstar.ca/Amstar_Checklist.php).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was an “a priori” design provided?</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes (update of older meta-analyses)</td>
<td>yes (update of older meta-analyses)</td>
<td>nein</td>
<td>ja</td>
<td></td>
</tr>
<tr>
<td>2. Was there duplicate study selection and data extraction?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>3. Was a comprehensive literature search performed?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>5. Was a list of studies (included and excluded) provided?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no (update of older meta-analyses)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6. Were the characteristics of the included studies provided?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>Yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>7. Was the scientific quality of the included studies assessed and documented?</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>Yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>8. Was the scientific quality of the included studies used appropriately in formulating conclusions?</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>Yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>9. Were the methods used to combine the findings of studies appropriate?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>Yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>10. Was the likelihood of publication bias assessed?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>No</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>11. Was the conflict of interest included?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
**Inclusion of primary studies**

It was also very noticeable that reviews often looked at very different primary studies. Although most reviews included the best known European RCTs (NTCC1 and 2, Swedescreen, ARTISTIC, POBASCAM, Finland study), the Indian study by Sankaranaryan et al. published in 2009 [21] and the Canadian CCCAST [22] were usually not included. The heterogeneity of the primary studies was also interpreted differently by different authors. This heterogeneity prompted some authors of reviews to dispense completely with any meta-analytic pooling of individual trials (e.g. Patanwala 2013 [9]). However, it was not possible to identify a link to the quality of the investigated meta-analyses determined using AMSTAR, e.g. that high-quality meta-analyses tend to be critical of the level of evidence.

**Results of the meta-analyses**

The different authors of the various meta-analyses also came to different conclusions and made different recommendations. To take one example, the joint study by Vesco (full report in 2011) and Whitlock (short summary in 2011), which served as the basis for the recommendations of the US Preventive Services Task Force, came to the conclusion that HPV testing using the Hybrid Capture 2 Test is significantly more sensitive than cytology only with respect to the detection of CIN 2+ (and of CIN 3+, less data available). Nevertheless the authors concluded that the available data was not yet complete, consistent, or relevant enough to permit a conclusive decision to be made about the role of HPV testing in primary screening.

In their interpretation of the data, the IQWIG 2011 and 2014 studies noted that the data appeared to show (Table 1) that the incidence of CIN 3+ and of invasive cervical cancer was lower with HPV screening and came to the conclusion that the level of evidence of a meant that HPV screening should be proposed to the German Federal Joint Committee as a general recommendation – although not as a strong recommendation [15, 16].

In their review and meta-analysis published in 2012 Murphy et al. recommended even more strongly that HPV testing should be introduced in the context of organized screening programs as the primary screening method for women over the age of 30 or 35 years. MaHTAS (Health Technology Assessment Section, Ministry of Health Malaysia) was also of the opinion that there was a high level of evidence showing HPV screening would be able to reduce the incidence and mortality rate of cervical cancer.

Overall, six of the 13 meta-analyses investigating this issue reported a clear benefit of using HPV test (potentially combined with cytology for co-testing) or recommended HPV test as the primary screening method (Arbyn 2015, Bouchard-Fortier, MaHTAS, MSAC, Murphy, and Pileggi). The remaining meta-analyses also reported benefits from HPV screening but were of the opinion that the data was not yet sufficient to support a general recommendation to use HPV testing to screen for cervical cancer. None of the analyses reported that cytology-based screenings offered more benefits compared to HPV screening.

**Discussion**

In Western countries all over the world prevention programs for cervical cancer are being revised. Currently, numerous countries (USA, Australia, Netherlands, Spain, Portugal, Sweden, Italy, Slovenia, etc.) are switching from previously established programs based on cytological screening to early detection of cervical pre-cancer based on HPV testing in women above the age of 30–35 years. In the USA recent publications by Ronco et al. [23] and the final results of the ATHENA study [24] have led to primary HPV screening being recommended since the beginning of 2015 as an alternative to cytological screening due to its equal or higher efficacy compared to cytology [25]. This paradigm change is based on data obtained from various high-quality RCTs (see above). When formulating recommendations, policymakers in the healthcare system base their decisions on the results of systematic reviews which summarize RCTs. Depending on the method and analysis used in these reviews, the reviews come to sometimes quite different conclusions – as was shown in this systematic summary.

Meta-analyses can arrive at different conclusions because they interpret the existing evidence differently. The IQWIG has highlighted the “very different screening strategies” and the “limited similarities between studies” [15]. Kleijnen also considers the heterogeneity of studies to be moderate to high and the level of evidence for this to be very low [17], and Patanwala entirely dispensed with meta-analytic pooling of individual studies because of their heterogeneity [9]. In contrast, MaHTAS in Malaysia [14] and the publication by Arbyn et al. in 2015 [18] reported that the level of evidence was good or strong, using the same data. The primary studies included in the reviews also differed significantly. The greatest differences between the different meta-analyses were reflected in the quality as measured using the AMSTAR criteria. Five reviews received only three or fewer points (out of a maximum of eleven) (Table 2); however, this may also be due to the fact that the documentation was not always very transparent (no description of the methodology used).

In the end, around half of all investigated meta-analyses found significant indications that screening based on HPV testing was superior to cytology or recommended HPV-based screening. The remaining studies considered that the data was not yet sufficient to warrant a clear recommendation. The aim of our study was to identify and present the impact of methodological groundwork on decisions taken by healthcare policymakers.

The Lancet article by Ronco et al. published in 2013 [23] was not included in our study. Their publication reviewed and followed up the data of around 180,000 women aged between 20 and 64 years studied in four RCTs (Swedescreen, POBASCAM, ARTISTIC and NTCC) over an average of six-and-a-half years. A total of 107 invasive cervical cancers occurred in that period; the ratio of the incidence rates was 0.60 (95% CI: 0.40–0.89) in favor of the HPV group. According to these data, the protection against invasive cancer was 60–70% higher with HPV-based screening. With regard to the screening intervals, the cumulative cancer incidence was lower five-and-a-half years after a negative HPV test than at three-and-a-half years after an unremarkable cytology-based test, meaning that screening based on testing for HPV every five years in women aged over 30 years promises greater safety than cytology screening carried out every three years [23]. HPV screening carried out every three years was not compared with annual cytology-based screening, as the latter is not done in four European countries and is not recommended by the IARC/WHO. It will be interesting to see which decisions will be taken with regard to future programs for the prevention of cervical cancer in Germany. In Germany, the Joint Federal Committee (G-BA) is responsible for establishing the guidelines. The S3 guideline “Prevention of Cervical Cancer” will be finalized shortly and will be published later this year.
Conclusion

Different meta-analyses which examined the same or similar issues came to quite different conclusions because of the different approaches used for analysis. This is quite significant because the primary data on which the meta-analyses were based did not change. The data indicate that HPV-based screening is superior. The heterogeneity of the meta-analyses affects the decisions taken by policymakers in the healthcare system.

Conflict of Interest

M. Jentschke received lecture fees and partial reimbursement of travel and conference expenses from Abbott Molecular, Wiesbaden. P. Hillemanns received research grants from GSK and Abbott and lecture fees from SPMSD, Roche and Hologic.

References

2 Marquardt K, Buttner HH, Broscheiwitz U et al. Persistent carcinoma in cervical cancer screening: non-participation is the most significant cause. Acta Cytol 2011; 55: 433–437
8 Pan QJ, Hu SY, Guo HQ et al. Liquid-based cytology and human papillomavirus testing: a pooled analysis using the data from 13 population-based cervical cancer screening studies from China. Gynecol Oncol 2014; 133: 172–179
24 Wright TC, Stoler MH, Behrens CM et al. Primary cervical cancer screening with human papillomavirus: end of study results from the ATHENA study using HPV as the first-line screening test. Gynecol Oncol 2015; 136: 189–197