The aftermath of a pill scare: regression to reassurance

Walter O. Spitzer

McGill University, Montreal, Canada

In October 1995, following confidential exchanges of findings among investigators in several epidemiological studies, the UK Medicines Control Agency sent a ‘Dear Doctor’ letter to all clinical practitioners in the country. The letter alerted them to the possibility of an excess risk of venous thromboembolism among women taking combined oral contraceptives (OC) with the ‘newer’ progestins, notably desogestrel and gestodene. The communication provoked a major pill scare, not just in the United Kingdom but in other countries. The preliminary and unpublished findings from the four initial 1995–96 studies reported odds ratios (OR) ranging from 1.5 to 2.3 in the point estimates. These are very low relative risks but were communicated in a way that the public perceived as a ‘doubling of the risks’. In the 3 years since the pill scare, additional research has been done. First, deliberate and careful analysis of some of the studies and replication of others have shown that the epidemiological investigations were affected by unavoidable systematic error. Three types of bias were demonstrated empirically, namely, prescription bias, referral bias and healthy user effect or attrition of susceptibles. All those biases would tend to drive OR spuriously upwards. Additional epidemiological studies have progressively shown lower ORs, some of them under the threshold of 1.0, i.e. ‘no association’. Two major consensus assessments, one carried out by a World Health Organization Scientific Group and another undertaken by the International Federation of Fertility Societies, both attach little importance to differences between older (second generation) combined OC and newer ones (third generation). This paper is a synthesis of all published evidence since October 1995, at the time of the pill scare and in the 3 years since. In conclusion, all combined oral contraceptive pills are equally safe.

Key words: cardiovascular disease/epidemiological studies/oral contraceptives/pill scare

TABLE OF CONTENTS

Introduction 736
The initial data 737
Studies since 1996 738
The serious cardiovascular outcomes 741
The Danish population-based comprehensive assessment of all risks 742
A comment on two late-breaking consensus reports 742
Did the feared epidemics occur? Were any epidemics averted? 743
Conclusions in 1995 and conclusions in 1998 744
The safety of oral contraceptives 744
References 744

Introduction

In the first week of October 1995, my colleagues and I shared the tentative results gathered by our Transnational Study on venous thromboembolism (VTE) in confidence with a subcommittee of the Committee on Safety of Medicines and senior medical officers of the Medicines Control Agency (CSM/MCA) (Spitzer et al., 1996). At the time, I was relieved to find that the odds ratios (OR) in our analyses were <2.0 when comparing third generation oral contraceptives (OC) with second generation OC as the reference.

Our research team immediately began the arduous task of verifying every calculation of an analysis that had been hastily implemented at the request of the CSM/MCA and the Bundesinstitut für Arzneimittel und Medizinische Produkte (BfArM) in Germany. We assembled a panel of very senior and independent academic leaders in biostatistics and epidemiology, moderated by Dr Stephen P.Lock, former Editor of the British Medical Journal, to give us an in-depth rigorous critique and to help us finalize plans for a definitive analysis.

Three weeks later, while the panel was still deliberating and advising our investigative team in Montreal, the CSM/MCA distributed a ‘Dear Doctor’ letter, which unleashed a major pill scare in Europe. Subsequent regulatory actions in several countries (notably Germany, the UK and Norway) resulted in a de-facto ban of third generation pills in late 1995 and early 1996.

1Address for correspondence: 15313 Isabella Court, Corpus Christi, TX 78418, USA
Table 1. Number of cases and controls using second or third generation oral contraceptives in case–control analyses on venous thromboembolism

<table>
<thead>
<tr>
<th>Study</th>
<th>Cases (n)</th>
<th>Controls (n)</th>
<th>Adjustment for duration of use</th>
<th>Matching for age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHO (1995)</td>
<td>88</td>
<td>124</td>
<td>None</td>
<td>5 year bands</td>
</tr>
<tr>
<td>Transnational: Spitzer et al. (1995)</td>
<td>259</td>
<td>651</td>
<td>None</td>
<td>5 year bands</td>
</tr>
<tr>
<td>GPRD: Jick et al. (1995)</td>
<td>75</td>
<td>300</td>
<td>None</td>
<td>2–5 year bands</td>
</tr>
<tr>
<td>LETS: Bloemenkamp et al. (1995)</td>
<td>64</td>
<td>44</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Newer data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danish case–control: Lidegaard et al. (1997)</td>
<td>151</td>
<td>178</td>
<td>Current use</td>
<td>None</td>
</tr>
<tr>
<td>MediPlus UK: Farmer et al. (1997)</td>
<td>85</td>
<td>313</td>
<td>Current use</td>
<td>Exact age</td>
</tr>
</tbody>
</table>

WHO = World Health Organization; GPRD = General Practice Research Database; LETS = Leiden Thrombosis Study.

Three years later, we have the chance to take a further look at these regulatory actions taken less than 3 weeks after the disclosure of data. We now have longitudinal perspectives that allow us to judge the appropriateness of the initial decisions, to examine the consequences of very rapid regulatory actions taken in late 1995, and to consider some lessons that the 1995 pill scare have taught us about pharmacovigilance, drug safety and public health.

In this synthesis of evidence, I plan to examine related clinical practice guidelines, as well as the implications for public health policy. I have subjected myself to the rule of confining the synthesis to published data in peer-reviewed journals. I shall first focus on the cluster of findings as they were eventually published in late 1995 and early 1996. I will give a pharmacoepidemiological interpretation of what those early results meant for advice on contraception and family planning. I shall then trace the major results that have been published since early 1996. This synthesis of evidence will also be presented within a theoretical framework for the elucidation of cause from an observed association. Finally, I will examine whether there is any evidence that populations of women of childbearing age have been affected by the use of newer products on the market or by decisions of regulators.

The initial data

The four studies published in December 1995 and January 1996 (Table 1) reported OR for VTE comparing third generation OC with second generation OC. The World Health Organization (WHO) and the Spitzer projects were case–control studies where new data were generated on the field following rigorous protocols which were very similar one to the other, by design. The Jick study was an analysis of existing data in the British General Practice Research Database (GPRD) and the Bloemenkamp study was a reanalysis of another study done in Leiden to investigate other questions. I only summarize the findings here since they have been systematically reviewed and interpreted in papers written since 1995 (Spitzer, 1997, 1998). The OR (relative risks) ranged from 1.5 to 2.3 (Bloemenkamp et al., 1995; Jick et al., 1995; WHO, 1995; Spitzer et al., 1996). At first glance the results appeared consistent, suggesting an increased risk of VTE when taking third generation OC in comparison to second generation OC. In the course of more definitive analyses, however, contradictions between the new data and earlier evidence came to light. The lack of a biologically plausible and clinically sensible explanation and the likelihood that systematic error (bias) played a major role became evident. Alternative interpretations now seem plausible and even probable.

The OR for VTE were consistent. It is noteworthy that they were consistently weak and that the lower limits of all the confidence intervals (CI) came close to or included 1.0. Odds ratios this small, even when statistically significant, are difficult to distinguish from ‘background noise’. Conventional epidemiological thinking as to OR hovering about 2.0, for a disease as rare as VTE, puts such risks below the threshold of public health concern. Indeed, for uncommon outcomes, I regard the ‘microscope of epidemiology’ to have insufficient resolution for relative risks <2.0.

If the OR is derived from case–control studies (and this was the case in all four studies published in December 1995 and January 1996), the unavoidable bias of methods used in case–control research compels even greater caution.

Problems in interpreting low relative risks

Historically, the reduction in the dose of ethinyl oestradiol in OC has been associated with a reduction in VTE incidence. OC containing 50 μg ethinyl oestradiol per tablet were associated with incidences of 7–11 per 10 000 woman-years, and OC containing 30–35 μg ethinyl oestradiol of ~4 per 10 000 woman-years (RCGP, 1974, 1978; Vessey, 1988; Gerstman et al., 1991; Farmer and Preston, 1995). Paradoxically, in the 1995–1996 studies, the OR for the OC containing 20 μg ethinyl oestradiol combined with desogestrel were higher, by a factor of 1.5–5, than those of the same progestogen combined with 30 μg ethinyl oestradiol (Jick et al., 1995; WHO 1995; Lewis et al., 1996a). In the WHO study, for cyproterone acetate-containing products, an inverse ethinyl oestradiol dose gradient
was found (WHO, 1995). These inverse gradients are clinically and biologically implausible and militate against a causal interpretation of the weak OR between third and second generation OC reported in the earlier studies.

Studies since 1996

Scientific evidence is never static. Whether the issues require explanatory research based on the fundamental sciences, or controlled clinical investigation on efficacy or observational research on safety, knowledge evolves. Even in the short span of 3 years since October 1995, a great deal of evidence has been generated and published in peer-reviewed journals. The ensemble of up-to-date evidence in late 1998 leads to different conclusions than those drawn from the evidence available in late 1995. In this next section, I summarize the findings of evolving research. Many of the details have been given in other contributions to this collection of papers. I shall highlight the pattern of the entire mosaic rather then give the detail of each individual analysis.

Studies bearing on systematic error

Age and duration of OC use are directly or indirectly related to a number of essential factors that affect the risk estimates of VTE. In addition, these factors are directly and strongly related to each other. Younger women are more likely to have used OC for a shorter period of time, for example. They are also more likely to be an OC starter compared to older women and they are more likely users of the newer OC on the market. Young women have, in addition, a lower chance of being a healthy long-term user of an OC with lower risk of VTE than the short-term users of the same OC. The direct relations of age and duration of OC use with both the exposure and the outcome in this particular epidemiological association corresponds with the definitions for confounding. It follows that it would be essential to allow the lowest possible biasing influence of these factors in the studies comparing OC use, either by exact matching of cases and controls, or by adjusting for the effects in the analyses.

By their nature, observational studies are vulnerable to systematic error or bias, and the results are less valid and less reliable than those from experimental randomized control trials. Randomised control trials evaluating the safety of OC have never been done and cannot be done for ethical, practical and methodological reasons. Theoretically, the epidemiological studies that were carried out and reported in 1995 and early 1996, since they were observational case-control studies, were expected to be affected by many biases. That is unavoidable in case-control research.

I shall only point to three types of bias because for these there is empirical evidence published since 1996 that these factors were at play in research done. The three biases are: prescribing bias, healthy user effect and referral bias.

Prescribing bias

Third generation OC have been preferentially prescribed to women with cardiovascular risk factors because of their perceived improved safety profile over second generation OC. The evidence for such is convincing. For example, a study in the UK, Germany, Sweden and The Netherlands found that, in the presence of cardiovascular risk factors, the likelihood of a prescription of a third generation OC was five times as high as that for a second generation OC (71% versus 14%, respectively) (Van Lunsen 1996). Similarly, a study in France showed that, of the 60% of gynaecologists and 52% of general practitioners who would actually prescribe an OC to women with a family history of thrombosis, 90% and 68% respectively would only prescribe a third generation OC (Jamin and De Mouzon, 1996). The impact of this pattern of practice on case-control study findings would have increased systematically and artificially the OR of third generation OC for VTE.

Healthy user effect or attrition of susceptibles

In my view, this bias alone could explain all the elevated OR observed between third and second generation OC for VTE in the 1995–1996 studies. Those users of drugs that are susceptible for (serious) side-effects will be partly or largely eliminated from the user pool by prudent clinicians over time. The clinicians would have recognized contraindications or side-effects in certain women, which may render them susceptible to VTE. This leaves the original cohort of women on second generation OC with fewer susceptible users. Since second generation OC had been available much longer than third generation, it is likely that the 1995–1996 studies were all affected by this bias.

The most compelling evidence for the presence of the attrition of susceptibles bias in the initial studies comes from an in-depth analysis of the Transnational study database, which revealed important time trends (Lewis et al., 1996a). The analysis showed that the product that had been on the market longest had the lowest OR for VTE and that each consecutive OC introduced had a higher OR for VTE than its predecessors. There was a clear correlation of the OR for VTE with the recency of market introduction of the products concerned ($P = 0.00012$, analysis for linear trend).

There is no known pharmacological rationale for this effect. I have referred to this type of systematic error in observational research as ‘play the winner’ bias. Others have used the phrase ‘prudent doctor bias’ in respect to other pharmaceutical products. What probably happens is that women doing well or free of a risk profile are kept on the older products that have served them well. The prudent doctor will not switch products for them and play the ‘winning’ OC. Higher risk women, in contrast, will be switched to newer products believed to be safer. The main effect of this bias in the 1995–1996 studies was to reduce the absolute and relative risks of VTE for
second generation OC (absolute in comparison with previous studies, relative in comparison with third generation OC).

Referral bias

Referral bias occurs if there is a selective tendency for ambulatory clinicians or primary care physicians practising family planning to refer women with suspected venous-thromboembolic conditions if they had been taking third generation OC. Conversely, the tendency would mean that fewer women on second generation OC would be referred to hospital for investigation in the presence of worrisome symptoms or findings. There is empirical evidence that this bias may have affected epidemiological studies comparing third and second generation OC. In interview surveys in Germany and in the UK, physicians indicated higher referral rates for third generation users in the presence of the same known or perceived risk factors for VTE (Heinemann et al., 1997; Dunn et al., 1998). Accordingly, third generation OC users would be more likely to contribute to the accrued hospitalised cases in all three epidemiological studies. It needs emphasizing that all the 1995–1996 case–control studies were hospital-based. Thus they are highly vulnerable to referral bias because women with symptoms suggestive of venous thrombosis and who use third generation OC are more likely to be referred for further diagnostics than women with similar symptoms using second generation OC, or not using OC at all. Adjustment for this bias in analysis is not feasible and prevention of this bias in fieldwork is virtually impossible.

The three categories of bias for which empirical evidence has been advanced is in the direction of increasing OR. Correcting them or allowing for them would reduce the initially reported low OR, which were already very low at ≤2.3.

Elimination of biases

To test the opinion expressed above, that correcting biases can and do reduce the initially derived OR, Suissa eliminated both the healthy user effect and duration of use as a confounder by analysing data in the Transnational database using only ‘OC starters’. This advanced analysis strictly controlled for duration of use as a covariate using logistic regression and quadratic spline models (Suissa et al., 1997). Equivalent adjusted OR comparing third and second generation OC showed no difference (OR = 1.0) when the analysis was restricted to starters and adjusted for duration of use (Suissa et al., 1997). There is considerable disagreement among statisticians about the smoothing mathematical techniques used by Suissa in the analysis. Nevertheless it is practically the only way to assess unstable numbers when the subset of subjects becomes very small when retaining only starters. I consider the findings corroborative of the general tendency to observe progressively lower OR as one controls better for confounders. The coherence of the trend towards lower risk ratios is the important observation; not the findings of any one study or any one analysis.

Adjusting for duration of use in a new database

Lidegaard, in a Danish case–control study involving 375 cases and 1041 controls, found that the OR of VTE among current OC users were influenced by duration of use (Lidegaard, 1997; Lidegaard et al., 1997a). Although no detailed information on OC use prior to the current OC was available in the database, the effect of adjustment for duration of use in the analysis was essential to the results. Without adjustment for duration of use there was a statistically significant difference in risk of VTE between third and second generation OC: OR 1.8 (1.0–3.0). The odds ratio decreased to 1.4 after adjustment for duration of use and was no longer statistically significant (0.8–2.5).

Effect modification

‘Effect modification occurs when a factor modifies the effect of a suspected causal factor under study. For example, age is an effect modifier for many conditions and immunisation status is an effect modifier for the consequence of exposure to pathogenic organisms’ (Last, 1988). All the initial studies attempted to take effect modification into account in the design of the studies. Accordingly, matching by age was done between cases and the corresponding controls for all patients accrued for the studies, either in fieldwork or from computerized databases. Several new studies and analyses have been completed and published since early 1996, which improved matching procedures to minimize effect modification (Farmer et al., 1998; Lawrenson and Farmer, 1998).

The UK MediPlus Study

This study (Farmer et al., 1997) involved ~80 OC-using women affected by VTE and had in essence the same architecture as one of the 1995–1996 initial studies, the Boston Collaborative Drug Surveillance Programme (BCDSP) study (Jick et al., 1995). The main difference between the designs is in the age-matching of the controls to the cases. In the BCDSP study, controls could be up to 2 years apart in age from the cases. In Farmer’s study, cases and controls had to have the same year of birth. The results show the sensitivity of the results to age-matching; the OR contrasting third and second generation OC were 0.8 (0.4–1.9) and 0.9 (0.4–1.8) for desogestrel- and gestodene-containing OC, respectively.

The German MediPlus Study

A similar study (Farmer et al., 1998) conducted with the German MediPlus database showed results similar to those in the UK. After matching cases and controls by exact year of birth, the unadjusted odds ratio for VTE of third generation OC users compared to second generation OC users was 0.8 (0.4–1.6).
Table II. Odds ratios (OR) for venous thromboembolism (VTE) (95% CI) third versus second generation oral contraceptives

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Reference group</th>
<th>OR for VTE (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gestodene</td>
<td>levonorgestrel</td>
<td>2.0 (0.6–4.7)</td>
</tr>
<tr>
<td>Transnational: Spitzer et al. (1996)</td>
<td>Desogestrel</td>
<td>2nd generation</td>
<td>1.5 (1.1–2.2)</td>
</tr>
<tr>
<td></td>
<td>Gestodene</td>
<td>2nd generation</td>
<td>1.5 (1.0–2.2)</td>
</tr>
<tr>
<td>LETS: Bloemaenkamp et al. (1995)</td>
<td>Desogestrel</td>
<td>levonorgestrel</td>
<td>2.2 (0.9–5.4)</td>
</tr>
<tr>
<td>GPRD: Jick et al. (1995)</td>
<td>Desogestrel</td>
<td>levonorgestrel</td>
<td>2.2 (1.1–4.4)</td>
</tr>
<tr>
<td></td>
<td>Gestodene</td>
<td>levonorgestrel</td>
<td>2.1 (1.0–4.4)</td>
</tr>
<tr>
<td>Newer data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danish case–control: Lidegaard et al. (1997)</td>
<td>3rd generation</td>
<td>2nd generation</td>
<td>1.4 (0.8–2.5)</td>
</tr>
<tr>
<td>MediPlus UK: Farmer et al. (1997)</td>
<td>Desogestrel</td>
<td>levonorgestrel</td>
<td>0.8 (0.4–1.9)</td>
</tr>
<tr>
<td></td>
<td>Gestodene</td>
<td>levonorgestrel</td>
<td>0.9 (0.4–1.8)</td>
</tr>
<tr>
<td>MediPlus Germany: Farmer et al. (1998)</td>
<td>3rd generation</td>
<td>2nd generation</td>
<td>0.8 (0.4–1.6)</td>
</tr>
<tr>
<td>Transnational: Suissa et al. (1999)</td>
<td>3rd generation</td>
<td>2nd generation</td>
<td>1.0</td>
</tr>
</tbody>
</table>

CI = confidence interval; WHO = World Health Organization; GPRD = General Practice Research Database; LETS = Leiden Thrombosis Study.

The British GPRD Study (Lawrenson and Farmer, 1998)

To evaluate the discrepancy in results between the initial BCDS study (Jick et al., 1995) and the UK MediPlus study (Farmer et al., 1997), and the effects of matching cases and controls by exact age, a further study was carried out using the same General Practice Research Database (GPRD) employed by Jick.

In a nested case–control analysis, 318 cases were matched by general practice and by exact year of birth to 1224 controls. The results showed no difference in OR for VTE between OC containing desogestrel, gestodene and levonorgestrel (see Table II).

Update on biological plausibility

Since 1995, communications to the public, to the health professions, to the press and through the lay press about the initial studies comparing second and third generation OC have neglected to put the initial findings in the context of biological plausibility. Until now, there has been no known biologically plausible mechanism that could explain the differences in the rates of VTE. A systematic evaluation of all the randomized control trials and basic research that investigated the haematological effects of second and third generation OC have not shown any difference that could be considered indicative of a higher (pro)thrombotic potential of third generation OC compared to the reference second generation products (Winkler, 1998).

The discussion on biological plausibility was revived in April 1997 by a paper that suggested the use of OC would lead to acquired activated protein C (APC) resistance, and that this outcome was even more pronounced with third generation OC (Rosset et al., 1997). The data on which these suggestions were based came from a cross-sectional study in which a new haematological assay to assess APC sensitivity with measurement of thrombin generation was performed with blood samples from different subgroups.

The study did not follow the broadly accepted design of metabolic studies and the use of the new assay generated more questions than answers. The study was not randomized, thus allowing selection biases to influence the results. In spite of the observed very large variability in APC sensitivity ratio during the monthly pill-taking cycle, no standardized timing of the blood sampling in the cycle was applied. The investigators pooled samples in one part of the investigation and assessed only the end concentration of thrombin in the other part of the study in order to save time. The assay has not yet been reproduced in other laboratories and it is unknown to what extent, if at all, the results are associated with the development of VTE, since they could not be reproduced by a validated APC sensitivity assay (Schramm and Heinemann, 1997). In short, evidence on biological plausibility is still lacking.
Table III. Diagnosing causality of oral contraceptives/venous thromboembolism studies

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>Not done and not doable</td>
</tr>
<tr>
<td>Strength</td>
<td>Very weak</td>
</tr>
<tr>
<td>Consistency</td>
<td>No longer present</td>
</tr>
<tr>
<td>Gradient</td>
<td>Reverse, paradoxical</td>
</tr>
<tr>
<td>Biological plausibility</td>
<td>Absent</td>
</tr>
<tr>
<td>Specificity</td>
<td>Lacking</td>
</tr>
<tr>
<td>Coherence</td>
<td>Missing</td>
</tr>
<tr>
<td>Temporality</td>
<td>Non-contributory</td>
</tr>
<tr>
<td>Analogy</td>
<td>Not found</td>
</tr>
</tbody>
</table>

**Bradford-Hill criteria for causality**

Odds ratios or relative risks are estimates of association. They are not proof of causality. In a landmark paper published in 1968, Sir Austin Bradford-Hill set forth a list of criteria or guidelines that have been used extensively in elucidating causality from association (Bradford-Hill, 1971). The categories of guidelines are given in Table III. I will assess the evidence of association of use of third generation OC with an excess of VTE against that framework.

No experiments bearing on the safety of any oral contraceptive have been done. As of the time of this publication, the range of relative risks comparing third generation OCs to second generation OCs is from 0.8 to 2.3. Odds ratios of about 2 and under are very weak when the outcomes are rare, such as VTE, or myocardial infarction (MI), or stroke in young women. At that low level, they have no clinical importance, nor any public health significance. Although the initial studies suggest a certain degree of consistency in OR for VTE when contrasting third versus second generation OC users, the more recent studies no longer show differences between the third and second generation OC. Consistency has been lost. Not only was there no expected gradient correlated with ethinyl oestradiol dose of combined oral contraceptives (COC) in all four studies that sought to demonstrate it (Jick et al., 1995; WHO 1995; Spitzer et al., 1996; Farmer et al., 1997); instead, a surprising reverse gradient was observed. Such a paradoxical finding argues strongly against causality.

Several earlier sections of this article have highlighted the lack of biological plausibility of an excess in the frequency of occurrence of VTE related to use of third generation OC. Moreover, there is no specificity in the associations observed, since VTE is known to be related to obesity, immobility, pregnancy and other risk factors. In addition, OC are not only associated with VTE. When the evidence from different disciplines and different sources 'hangs well together', one considers it to be coherent. Yet there are discrepancies in rates of occurrence of VTE between recent and earlier studies where those rates can be calculated for second generation OC users. The incidence of VTE in users of second generation OC in most recent studies is 50% lower than the incidences found 5-10 years earlier (Vessey, 1988; Gerstman et al., 1991; Jick et al., 1995; WHO, 1995). The paradoxical gradient does not contribute to coherence. The lack of support in laboratory research also mitigates against coherence. Temporality does not contribute to the assessment of causality because the studies were designed to assure that the exposure to OC preceded target events. I have not found any studies that permit reasoning by analogy using similar associations or similar causal relationships.

Table III summarizes the review of the evidence on causality. None of the guidelines or criteria appears to be met. The profile of the associations observed does not even remotely suggest causal relationships.

**The serious cardiovascular outcomes**

During the years I practised clinically and advised women on OC, the possibility of stroke among women taking the pill was my biggest worry. Stroke is the only serious adverse event I actually encountered with my patients, whether on first or second generation OC. All of us in primary care practice also worried about fatal acute myocardial infarction (AMI) and about blood clots in deep veins; but stroke, with its ensuing life-long major disabilities was the chief concern. VTE is a minor disease, easily treated on an ambulatory basis, and with a very low case fatality rate, usually reported to be <2%. Case fatalities for MI approach 50%, with half of those fatal cases never reaching the hospital. Case fatality rates for stroke are reported to be from 15% to 25%. Fortunately, recent studies provide reassuring results.

**Stroke: evidence that absolute and relative rates are diminishing**

The risk of stroke is gradually decreasing from the earliest epidemiological findings in the 1960s and 1970s and the absolute rates of occurrence of stroke have been declining steeply (Petitti et al., 1996; Heinemann et al., 1997; Lidgard, 1998). The published OR comparing second and third generation OC in respect to thrombotic stroke have revealed no differences in risk between the two types of products (WHO, 1996a; Heinemann et al., 1997). I have found no reports of differences for haemorrhagic stroke (WHO, 1996b).

**Acute myocardial infarction: possibly a protective role of third generation OC**

Studies published since 1996 focus on relevant cardiovascular disorders with higher burdens of suffering to those affected than VTE. They show consistently favourable results for third generation OC in respect to AMI compared to second generation OC (WHO, 1997; Jick et al., 1996). The final results from the planned analysis of the Transnational study show that the OR comparing third and second generation OC
was 0.3 (0.1–0.9), a two-thirds reduction of AMI among users of third generation compared to second generation OC (Lewis et al., 1997). Although the Transnational was the only study large enough to have power to attain statistical significance, all other studies show trends in the same direction (Jick et al., 1996; WHO, 1997). The worst case scenario is that third generation OC convey no risk in comparison with ‘no use’. When compared with second generation OC, it is likely that protection occurs when taking third generation OC.

The Danish population-based comprehensive assessment of all risks

One of the latest series of studies has been part of a Danish project led by Lidegaard. The distinctive feature of the Danish studies is that the database employed covers the entire population of the country, including the entire population of women of childbearing age (Lidegaard et al., 1997). Accordingly, selection bias has been eliminated. Because cases and controls can be shown to be probability samples of the entire population, it is possible to derive absolute risks. Another important, distinctive feature is that the databases contain information on case fatality, on background incidence of all three main categories of cardiovascular and cerebrovascular adverse events (VTE, AMI and stroke), as well as information on major disability.

Lidegaard’s group factored in rate of occurrence of all of these events, mortality, case fatality rates, residual disability, as well as duration of use. It was then possible to estimate the attributable risks for third generation OC or for second generation OC. Lidegaard was able to establish that if the entire population of Danish women of childbearing age were taking second generation OC, they would have 180% higher mortality, 30% higher morbidity and 220% higher disability than if the entire population were taking third generation OC (Lidegaard et al., 1997). If the Danes had evaluated VTE in isolation, this perspective would not have been detectable.

Balanced pharmacovigilance requires simultaneous assessment of all probable categories of adverse events and a consideration of incidence, morbidity, case fatality and residual disability. Whenever possible, such evaluations should be carried out on entire populations, or on probability samples of those populations, as was feasible in the Danish programme.

The data from their own governmental programme suggests that Danish women as a whole would be better off on third generation OC.

A comment on two late-breaking consensus reports

This review and synthesis has had as its aim to evaluate only original work published in peer-reviewed journals, including the most recent articles. However, the rule can and should be broken conservatively. As this review was being prepared, two very important consensus reports were published, one sponsored and carried out by the World Health Organization (WHO, 1998) and the other one by the International Federation of Fertility Societies (IFFS, 1999). Both working parties included authoritative experts from the clinical world, from the laboratory and from methodological centres for epidemiology and biostatistics. The WHO group involved many investigators who had actually conducted the studies in the preceding years. The IFFS group heard views from the investigators but those formulating the consensus were only those hitherto uninvolved in the conduct or analysis of any relevant study. It is significant that major convergence of opinion resulted in the two reports. These reports cannot be ignored in summarizing the findings on OC over the last 3 years.

The WHO Technical Report 877: cardiovascular disease in steroid hormone contraception

This report of a WHO Scientific Group that includes many scientists who were involved as investigators in the WHO study (WHO, 1995) and the Transnational Study (Spitzer, 1996) is well written, well reasoned and I agree with most of what it contains. My only substantive objection is with one conclusion found at the end of section 7 (page 50), which reads:

Combined oral contraceptives containing desogestrel or gestodene probably carry a small risk of venous thromboembolism beyond what attributable to combined oral contraceptives containing levonorgestrel.

In contrast, I have come to the conclusion, as documented and justified in this article, that OC containing desogestrel and gestodene do not carry even a small risk of VTE beyond that attributable to COC containing levonorgestrel.

My disagreement with the Scientific Group that wrote the Technical Report series occurs because they set aside clear and sometimes replicated empirical evidence about a series of biases, which I have summarized in earlier sections of this review. All these examples of systematic error would elevate OR spuriously in the early studies. For instance, Heinemann’s demonstration of referral bias (Heinemann et al., 1996) was replicated by Dunn (Dunn et al., 1998) in the UK. Further, prescribing bias is likely to have played a major role in driving OR higher. According to a large number of studies, analyses and surveys, third generation OC have been selectively prescribed to women according to the following user characteristics: starters (short duration of use), age, personal history of VTE, family history of VTE, clinical venous signs, chronic inflammatory disease, anaesthesia or plaster cast, obesity, diabetes, standing working position, alcohol abuse, smoking or a combination of risk factors (Lis et al., 1993; Heinemann et al., 1996; Jamin and De Mouzon, 1996; Lewis et al., 1996a; Poulter et al., 1996; Van Lunsen, 1996;
Farmer et al., 1997; Lidegaard et al., 1997; Dunn et al., 1998).

Referral bias and diagnostic bias are often confused. There is a great deal of evidence that such bias occurred in the countries where the Transnational Study was carried out. Evidence of healthy user effect bias (Lewis et al., 1996a) and the distinctive experience of starters of OC is dismissed readily (Suissa et al., 1997). I find that surprising given that the bane of observational research, especially case–control studies, is bias.

I am more concerned by what was not written in the WHO Technical Series Report. Even if one were to accept that desogestrel or gestodene carry a small risk of VTE beyond that attributable to COC containing levonorgestrel, there is insufficient emphasis in the Technical Report upon the fact that derivation of such low relative risks to absolute risks of serious sequelae including death, are infinitesimal, of no clinical importance and definitely of no public health significance. In my view, due to an omission, the WHO Technical Report is vulnerable to misinterpretation on the part of clinicians and patients. Moreover, an assessment of VTE in an isolated manner disregards the importance of 'no difference' in occurrence rates of MI between newer individual products and the older products and 'no difference' in stroke. Myocardial infarction and stroke, as discussed in earlier sections, are the vascular adverse events of serious concern among young women in the prescription of OC.

I fully agree with all the final global conclusions of the WHO Scientific Group, notably the following:

(i) The incidence and mortality rates of all cardiovascular diseases (stroke, AMI and venous thromboembolic disease) in women of reproductive age are very low.

(ii) Any increase in incidence of or mortality from cardiovascular disease attributable to use of combined oral contraception is very small if users do not smoke and do not have other risk factors for cardiovascular disease.

(iii) At any given age, a woman who smokes but who does not use OC is at greater risk of death from arterial disease than a user of OC who does not smoke.

(iv) Venous thromboembolic disease is the most common cardiovascular event among users of OC. However, it contributes very little to any increase in the number of deaths since the associated mortality is relatively low compared to that associated with arterial diseases. Long-term disability from non-fatal thromboembolic disease is also low.

International Federation of Fertility Societies Consensus Conference

I was surprised that the text of the Consensus Report appears to give so much weight to the findings of the WHO Study (WHO, 1995) suggesting that it is very large: 1143 cases and 2998 controls primarily hospitalized women from 21 centres around the world. In fact, the relevant findings about different types of OC ('old' and 'new') are drawn primarily from just one centre in Oxford and there are just 88 cases and 124 controls (Table IV). Otherwise I have no disagreement with the entire text of the document. The chief conclusion is:

Table IV. Number of cases and controls using second or third generation oral contraceptives in case–control analyses on venous thromboembolism

<table>
<thead>
<tr>
<th>Study</th>
<th>Cases (n)</th>
<th>Controls (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO study (1995)</td>
<td>88</td>
<td>124</td>
</tr>
<tr>
<td>LETS: Bloemaenkamp et al. (1995)</td>
<td>64</td>
<td>44</td>
</tr>
<tr>
<td>GPRD: Jick et al. (1995)</td>
<td>75</td>
<td>300</td>
</tr>
<tr>
<td>Transnational: Spitzer et al. (1995)</td>
<td>259</td>
<td>651</td>
</tr>
<tr>
<td>Danish case–control: Lidegaard et al. (1997)</td>
<td>151</td>
<td>178</td>
</tr>
<tr>
<td>MediPlus UK: Farmer et al. (1997)</td>
<td>85</td>
<td>313</td>
</tr>
<tr>
<td>MediPlus Germany: Farmer et al. (1998)</td>
<td>42</td>
<td>168</td>
</tr>
<tr>
<td>Transnational: Suissa et al. (1999)</td>
<td>65</td>
<td>171</td>
</tr>
<tr>
<td>(first time users)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPRD: Farmer et al. (1999)</td>
<td>335</td>
<td>1114</td>
</tr>
</tbody>
</table>

WHO = World Health Organization; GPRD = General Practice Research Database; LETS = Leiden Thrombosis Study.

The evaluating committee of the IFFS finds no reason to advise selective prescribing of OC containing different progestins on the basis of their effects on CVD [cardiovascular disease].

Their statements on the appropriate selection of candidates for OC had been advocated earlier by the WHO investigators (WHO, 1996). In the past, I have adopted public positions that are similar, supporting the notion that the armamentarium of safe OC currently approved by regulators and on the market should not be reduced (Spitzer, 1997). Further, judicious tailoring of the appropriate OC for each woman by skilful clinicians is the principal lesson learned from epidemiological research during the last 3 years.

Screening for genetic defects

I strongly agree with conclusions reached by the writers of the WHO Technical Series Report (WHO, 1998) and the IFFS (IFFS, 1999) that screening for genetic defects or for non-clinical markers of disease is inappropriate. Because of the low positive predictive value of those markers, one would risk denying OC to many women who could safely take them had they not been inappropriately labelled as 'high risk' by such screening.

Did the feared epidemics occur? Were any epidemics averted?

Three years after the ‘Dear Doctor’ letter was sent out to British clinicians, it is a good time to ask such questions. As documented in an earlier article (Spitzer, 1997), descriptive
epidemiology of VTE, AMI and stroke reveals that we have only observed non-epidemics in respect to all those disorders. But there have been epidemics of abortions (Office for National Statistics, 1997; Skjeldestad, 1997) in the UK, Norway and in Germany. The dire warnings issued by regulatory agencies have been shown to be false alarms. But the 1995 Great Pill Scare has taken its toll in other ways.

Conclusions in 1995 and conclusions in 1998

The range of OR found in the initial studies led me to conclude in 1995, as the findings were emerging from analyses and published, that the risks associated with third generation OC were slightly higher than those pertaining to second generation OC. My conclusions were based on rapid analyses demanded by regulatory agencies in late 1995. Subsequent careful analyses of my own Transnational study conducted at a measured pace brought to light biases and new perspectives that have forced me to re-examine my initial conclusions. Upon reflection, it became clear that a rate ratio contrasting third generation OC and second generation OC would be determined by the extremely low absolute rates of risk that were attained by second generation OC. Those very low rates were the reference product in most studies. Successive subsequent studies mandated by the CPMP (Committee for Proprietary Medicinal Products) in the EMEA (European Agency for the Evaluation of Medicinal Products) gave empirical evidence of biases that would all tend to drive OR artificially upwards.

Further studies progressively provided lower OR, some of them <1.0. A systematic appraisal of causality within the framework of Bradford-Hill guidelines does not suggest that any apparent excess of risk of third generation OC is the result of a cause and effect relationship. Indeed, a synthesis of all the findings from studies bearing on bias leads me to the conclusion today that many differences observed in 1995 are more likely due to bias than to cause.

In respect to VTE, my opinion has evolved as the profile of evidence has evolved. Today I am of the opinion that there is no difference between the risk profiles of third generation and second generation OC. An integrated balanced analysis, taking into account all the cardiovascular and cerebrovascular adverse events simultaneously, favours the view that the newer third generation OC are less risky to an unselected population of women of child-bearing age than the older OC on the market.

The safety of oral contraceptives

In terms of absolute rates of occurrence of cardiovascular events, the pill has been safe since it was introduced. In absolute terms, the cardiovascular risks have been real but always very low. There has been a progressive reduction of its risks with the reduction of the dose of the ethinyl oestradiol component of COC. The reductions have been particularly steep in respect to stroke.

For women, the pill is the second most important health protective intervention in the history of public health, second only to the prevention and control of communicable diseases. Anything that undermines the credibility of the pill threatens the health and the quality of life of millions of women and their families. The message I take from the full body of all the evidence that has been published in the last 3 years is that all pills on the market are very safe and that they are becoming progressively safer.

References


Safety of oral contraception  745


