Conclusions: the relative safety of modern oral contraceptives

Jean Cohen\(^1\) and R.G.Edwards\(^2,3\)

\(^1\)Hopital de Sevres, 8 Rue Marignan, Paris, France, and \(^2\)Editorial Office, ESHRE Journals, Madingley Road, Coton, Cambridge CB3 7PG, UK

Published data on the relative safety of second and third generation oral contraceptives (OCs) is critically assessed. The original four studies published in 1995/1996 and their conclusions are examined in detail, including comments made by other investigators. Each study has advantages and flaws which are balanced in detail. Newer studies are then examined in the same manner. All recent papers indicate an equivalence of safety of second and third generation OCs, as the effects of various confounders and biases have been identified and analysed. We conclude that absolute risks are minor with both generations and that no reasons now exist to indicate any differences in their overall safety. We analyse finally how the pill crisis arose, and make suggestions about future reporting on OC safety.

Key words: myocardial infarction/oral contraception/pill scare/safety/venous thromboembolism

TABLE OF CONTENTS

- Epidemiological debates and conflicts on the safety of oral contraception 756
- Deciding on the relative safety of different OC formulations 764
- How the pill crisis arose 766
- Conclusion: The essential background to the pill scare 769

Introduction

In this paper, we wish to draw our conclusions on the current safety of modern oral contraception. Before doing so, we must assess the epidemiological evidence, and clinical data to a lesser extent, as reviewed in our Introduction (Edwards and Cohen, 1999) and in successive articles of this publication. We also wish to clarify the events and causes of the pill scare which triggered widespread public anxiety. Finally, we make some suggestions how the catalogue of events should have been handled better by regulatory authorities, and how such eruptions may be avoided in the future regulation of oral contraception.

We remark initially that it is a lamentable but perhaps inevitable fact that wide differences still exist among epidemiologists about the relative safety of various oral contraceptive (OC) formulations. These controversies are discussed and judged in detail below. The final sections of our paper may suffice for readers less interested in such epidemiological detail.

\(^5\)To whom correspondence should be addressed

Epidemiological debates and conflicts on the safety of oral contraception

The nature of epidemiology in relation to OC safety

Since the pill scare, a great deal of epidemiological information including several large-scale studies emerged on the safety of OCs. Safety is a matter for detailed and sophisticated epidemiological analyses as described in our introductory paper (Edwards and Cohen, 1999). Epidemiology is the only means of assessing safety concerns in biomedicine, despite its inherent limitations, and epidemiologists differ over their interpretations of the various studies on safety of OCs. There can be no doubt about the high quality of the numerous studies applied to the safety of second and third generation OCs, yet virtually every published study has been criticized, then defended by its authors. We must therefore attempt to balance strengths and weaknesses of each study including the most recently published data.

Epidemiological projects are large undertakings, often involving years of preparation and data collection and then long periods of analysis. Immense care is needed in planning the strategy of the study, coping with physical and behavioural aspects of cases and controls, clarifying associations and — hopefully — causes, maintaining high standards in field work, and applying full statistical analyses to rare clinical events in the assembled data. Some causes of bias, e.g. body mass index (BMI), smoking, pregnancies, family histories, and the effects of modifying factors such as age and duration of use can be
measured cleanly. Even the most careful analysis may fail to measure complex social aspects of human behaviour affecting the investigation, although some effects of personal and social factors are increasingly recognised in recent studies, e.g. the duration of use or the switching of contraceptives. Decisions by clinicians, each with their own standard of care, may also have unmeasured consequences, to cause diagnostic or prescribing bias. Weaknesses in studies are exposed when published data are discussed and reinterpreted, or during critical discussions among epidemiologists.

Occasionally, wide agreement is reached between several investigating teams over specific issues. A striking example concerns venous thromboembolism (VTE) risk and oestrogen levels in various formulations. Three of the initial studies reported how OCs with 20 μg ethinylestradiol have a higher odds ratio for VTE than those containing 30 μg. The Transnational study reached the same conclusion for women aged 25–44 (Lewis et al., 1996a). Several investigating teams questioned their own similar conclusion in the absence of any biological justification, and so ascribed it to unidentified bias or confounding such as switching to OCs containing low concentrations of oestrogens by some high risk VTE users. Vandenbroucke et al. (1997) suggest that the higher risk ratio arose through a combination of starter and third generation effects. Grant (1997) prefers to ascribe it to differential sensitivities of physiological systems, each displaying its own unique peak effect in relation to oestrogen/progestosterone balance. Some varicocities for example fail to improve immediately on discontinuation of OC use in contrast to other conditions such as cramps. Endocrinologists propose that estrogenic effects of ethinylestradiol are enhanced by the lipid-favourable and less androgenic progestagens in third generation OCs (C.Jamin, personal communication). Whatever the explanation, such epidemiological effects are probably too small to seriously influence odds ratios.

A second form of agreement recently emerged over the interpretation of epidemiological data in relation to social factors, user behaviour and data retrieval. Jick et al. (1998b) analysed UK General Practitioners Research Database (GPRD) data after the pill scare which indicated no subsequent effects on conception or abortion among the population. They suspected their own conclusion and suggested that many women had by-passed their GPs when seeking advice on contraception and abortion. This bias was clarified by Puredi and Paintin (1998) who pointed out how the UK Office of National Statistics (1999) revealed 26 000 more recorded conceptions in 1995, and >13 000 abortions in 1996. This discrepancy with GPRD data probably arose when 3 million UK women annually preferred family planning services rather than consulting their GPs when deciding on matters of contraception, so their data never entered the GPRD database.

To set the current scene on the safety of OC formulations, we now briefly recapitulate and assess some of the earlier studies and controversies that characterize the past 5 years. Full details of these papers were presented in our Introduction to this symposium (Edwards and Cohen, 1999). We will then consider some very recent studies published within the last 2 years.

**GPRD and Leiden studies of 1995**

Clear disagreements have characterized epidemiological analyses since the onset of the pill scare in 1995. Among the four original studies on the risks of VTE, acute myocardial infarction (AMI) and stroke in relation to the use of OCs, two were modest in size, namely the GPRD analysis of Jick et al. (1995) and the Leiden study of Bloemenkamp et al. (1995). Two others were large and very detailed and closely similar in their design, namely the World Health Organization (WHO) and Transnational studies (WHO, 1995a,b; Spitzer et al., 1996).

Each study has been criticised as described in detail in our Introduction (Edwards and Cohen, 1999). Cramer (1996) was amongst those criticising Jick et al. (1995) on the grounds that the authors had underascertained cases of deep vein thrombosis (DVT), a comment denied by Poultet et al. (1996). Lidegaard et al. (1999) comment that the statistical power of the study was too weak to adjust for length of OC use. D.E.Walters (personal communication) points out that removing two positive VTE cases would destroy the statistical significance of Jick’s study, even though almost 250 000 patients were sampled. This step seems unnecessary since the diagnosis of this condition in the study was robust. The incidence of VTE in this analysis was much lower than in other large scale studies, and Farmer et al. (1999) question this low value and request explanations. Jick et al. (1995) stress how their lower values arose because they took care to include only well-diagnosed cases.

Bloemenkamp et al. (1995) were concerned largely with factor V Leiden. Their paper actually reanalysed data taken from another published study which was primarily an unmatched analysis concerned with the effects of factor V Leiden on VTE risk. Its data were heterogeneous and the study was not designed to compare the relative safety of second versus third generation OCs. They also met criticism from Cramer (1996) on conflicts between their two published papers on this topic. Some users were excluded from the second part of the analysis, a step that could have influenced its results and interpretation. Moreover, this was an unmatched analysis and many pairs were not intact in its database (Farmer et al. 1999). Nor did it assess the effects of duration of use.

**WHO studies**

The WHO study had large databases, skilful analyses and covered many countries. Each disease linked to OC use was assessed in detail (WHO, 1995a,b, 1996a,b, 1997, 1998). Its...
original aims included assessments of risks of DVT, pulmonary embolism (PE) or both among current OC users in nations worldwide. Secondary aims included the overall comparative safety of second and third generation OCs, among other factors. Among the initial studies and perhaps the major trigger of the pill scare in 1995, it identified relative risks of VTE of 2.6 for gestodene and desogestrel OCs versus levonorgestrel OCs. The significance of factors such as duration of use and OC switching in these studies were apparently not as fully appreciated at that time as they are today, and insufficient numbers of third generation OC users were recruited in most countries. Even today, the WHO study still paints the most critical analyses on VTE and third generation OCs and its conclusions are still widely quoted. Apparently a large study with >1000 cases and >2000 controls, WHO data on the risk of VTE with third versus second generation OCs was smaller than other studies and drawn mainly from Oxford with 80 cases and 124 controls (Spitzer, 1999). The Transnational study and some GPRD analyses are larger, with >250 cases and >600 controls.

WHO investigators are confident of the robustness of their data, and refer to it when criticizing other studies. Their criticisms sometimes rebound, encouraging counter-charges against the WHO study itself. One such response, by Lidegaard and Milsom (1996), in answer to charges against their own work by Farley et al. (1996), pointed out how a lack of control over bias and confounding characterize virtually every study on VTE risk in users of OCs, and listed important confounders including age, family history of VTE, BMI, varicose veins, previous births, hypertension during pregnancy, duration of OC use and referral practice. They adjudged the WHO study had failed to control adequately for age and BMI, insufficiently for varicose veins and hypertension during pregnancy (due to missing information), and not at all for family history of VTE and referral practice. Preferential prescribing could have arisen. Lidegaard and Milsom (1996) stress the risk of bias in the WHO study of VTE by its comparison of cases mostly against hospitalised controls. Community controls were available only for Oxford and then displayed differential responses among different age groups (Poulter et al., 1996). Comparisons based on community controls were accordingly quoted less frequently than those using hospital controls, although they gave adjusted relative risk for gestodene OCs which were lower than for levonorgestrel OCs (WHO, 1995b). Lidegaard and Milsom (1996) suggest the use of community controls by Farmer (1996) also produced lower odds ratios for VTE risk than in previous studies. Lidegaard and Milsom pointed out how the similar numbers of deaths arising in users of both second and third formulations – and deaths are not susceptible to referral bias – is consistent with the existence of bias for VTE risk in the original four reports. They also comment that healthy user effects identified in several other analyses must have been present in WHO data especially in relation to lower risks with levonorgestrel products. They also wished to know why WHO reported no effect of duration of use in one report (WHO 1995a) and clear indications of it in another (Poulter et al., 1996). They asked WHO to explain their higher VTE risk with low-oestrogen formulations, and asserted that selection bias due to prescribing bias and recency of use offered an explanation. Nor did WHO report on relationships between market introduction and VTE risk.

Poulter et al. (1996) defended the WHO analysis. Accepting the possibility of some bias in relation to validity of diagnosis when interpreting VTE and PE risks in users versus non-users, they pointed out how it was less likely to account for differences between OC types. Poulter et al. (1996) also stressed the consistency of the DVT:PE ratio in the WHO study since it resembled the expected 1:1 as found by Jick et al. (1995). They also believed that WHO use of hospital controls alone was justified by the study of Jick et al. (1995), whose community based controls produced risk estimates similar to WHO hospital controls, and by a need to avoid unreliable information from community controls. Nor could they accept that preferential prescribing was important in the WHO study, and it could not explain the excess risk of third generation OCs or its persistence into later years of use.

Lewis (1999) in the present publication in turn criticises Poulter et al. (1996). Their comparison of first versus previous users compounds problems of healthy user effects within groups of previous OC users and could produce higher estimates for newer OCs. The study design was violated by using a matched analysis even though the study was not matched for first users. The strength and analyses of the WHO database and its analyses were also criticised by Suissa and Spitzer (1998), while responding to criticisms made by Farley et al. (1998a) of their use of spline conversions of Transnational data. Suissa and Spitzer (1998) pointed out how abstractions of WHO data for similar spline conversions for duration of use reduced numbers to astonishingly few cases and controls, indeed to only 20 VTE cases and 12 controls which were highly inadequate for such complex transformations. Numbers were so few that spline conversions of WHO data by Farley et al. (1998a) apparently had to be truncated.

Heinemann, who helped with others to design the study, criticizes especially the WHO paper devoted to safety of second versus third generations OCs (WHO, 1995b; Heinemann, 1999). He stresses the initial intent was not to study differences between second and third generation OCs as a primary focus. Nor did he approve of Walker's (1998a) analyses of non-WHO studies with its contention that WHO studies were 'robust'. He also comments how the major source of VTE data came from Oxford, the only centre where relevant hospital controls accompanied the use of community controls, and how relative risks of gestodene versus levonorgestrel OCs adjusted for BMI and smoking were similar with community controls.
In answer to these criticisms, it must be stressed that the authors, and retrospectively the WHO Scientific Group (WHO Technical Report, 1998), strongly defended their studies. They believed that hospital based controls alone offered an adequate basis to estimate OC risks among users in the WHO study. When data were examined on cases where both controls existed, the WHO Scientific Group wrote 'they found similar relative risks associated with the use of oral contraceptives containing desogestrel or gestodene, regardless of which control group was used in the analysis'. Data given on community controls in the original report (WHO, 1995b) shows that only 18 cases were affected in this way, including 20% of the gestodene cases, 10% for levonorgestrel and 4% for desogestrel. The WHO Scientific Group thus presented an extensive discussion explaining why community controls were less satisfactory than hospital controls even though after adjustments the relative risk for gestodene versus levonorgestrel OCs declined to 0.9 whereas that for desogestrel OCs remained at 1.8. Yet, if community controls can be written off on the basis of 18 cases lacking controls, why were they collected in the first place? Surely, they are a fundamental element in any epidemiological assessment, perhaps especially in studies on safety of OCs, where many controls and cases with VTE may never enter hospital. Each control counterbalances the other. Indeed, the WHO report (1995b) itself states that 'in areas where there is homogeneous coverage of health services such (community) controls may be more representative of the population from which the cases arise'. Surely this condition exactly fits the Oxford area.

**Transnational study**

This was a very large study on the safety of second and third generation OCs, designed specifically for this purpose (Spitzer et al., 1996). Its analyses have been, and still are, being published in a series of ongoing studies. Among its 471 cases and 1,772 controls, a relevant 259 cases were matched with 651 controls from hospitals and community for purposes of analysing second and third generation OC users. Its original findings on VTE confirmed a higher relative VTE risk with third generation OCs, with an odds ratio of 1.5, i.e. lower than that estimated by the WHO analyses. The original Transnational study had to be published hurriedly in response to the pill scare. Successive studies originally planned when it was designed clarified bias and confounding in the original paper, and began a trend in a sequence of studies indicating that odds ratios for second and third generation OCs differed very little if at all.

This clarification began when adjusting for longer exposure times was found to reduce odd ratios to 1.4. It continued when duration of OC use was reassessed in detail in relation to the attrition of susceptibles or healthy user effect (Lewis et al., 1996a), to suggest that differential VTE risks among users of second and third generation OCs were even lower. A significant linear trend emerged when odds ratios for women aged 25–44 grouped in 5 year age bands were arrayed against date of introduction of various OCs to the market. This step led to conclusions that VTE risk peaked with recently-marketed OCs and then declined due to attrition of susceptibles. Lewis and Spitzer (1997) produced further evidence showing how more users of OCs for <2 years chose gestodene OCs rather than levonorgestrel OCs, while the latter predominated after 2 years. A group of healthy users had therefore emerged among levonorgestrel users. Further reanalyses by Suissa et al. (1997) utilising spline conversions revealed an equality of risk for VTE over time when duration of use and first-time uninterrupted use of second and third generation OCs were compared. Moreover, an unexpected benefit of third generation OCs emerged when the Transnational analysis on AMI and OC use was decisive in revealing an enhanced safety in some users of third generation formulations versus controls. Odds ratios for AMI among some categories of third versus second generation OC users thus declined to 0.28 or even less (Lewis et al., 1996b, 1997).

The authors pointed to other studies displaying similar trends for myocardial infarction (MI) risk (Jick et al., 1996; Lidegaard and Edstrom, 1996; WHO, 1997), even though only the Transnational analysis was significant because of its large size. The Transnational study thereby presented an impression far more reassuring on the safety of third generation OCs for VTE and MI than that given by WHO.

The Transnational analysis attracted many critics. Vandenburgroos et al. (1996, 1997) stressed the danger of overdoing re-analyses which might exploit spurious associations and reduce estimates of bias to null. They agreed with Weiss (1997) in objecting to the removal of data on VTE risk in women aged 18–24 years by Lewis et al. (1996a) and so questioned the claim on duration of use and attrition of susceptibles. Several critics queried the successive estimates of VTE risk as new sections of the Transnational study were published. The WHO Technical Report (1998) commented how the low MI risk reported by Lewis et al. (1996a) must be checked since very few cases used third generation OCs, and no previous study had identified such an effect. Moreover, this protective value of third generation pills actually emerged only in countries of continental Europe which may have different prescribing methods from UK. D.E. Walters (personal communication) also points out that only six cases of AMI were involved in the study of Lewis et al. (1997), with another one added in a later paper. Each of these charges was answered by Lewis and Spitzer (1997), who pointed out that healthy user effect can only be identified among users who had used OCs for several years, so that larger numbers of longer-term users used second generation OCs as each new user cohort stabilised its choice of OCs. Spitzer (1999) commented that the Transnational study was large enough for this study on MI, with other studies showing similar trends.
Strong opposition emerged against spline conversions of Transnational data to illustrate a similarity in risk of VTE with third and second generation OCs (Suissa et al., 1997; Suissa and Spitzer, 1998). Farley et al. (1998a) questioned the model as distorting original data, and failing to explain similar conversions of WHO data, so its generality was questionable. They felt that the model was inappropriate, mis-specified by constraining the splines to start at a risk ratio of 1 (as in non-users) and preferred an unconstrained spline model. This constraint applied by Suissa et al. (1997) prevented the model from reflecting the sharply increased VTE risk in users during their first months and an unconstrained spline would have given a closer fit to the original data. Walker (1998a) also questioned the spline modification when comparing different generations of OCs. He wished to see more details of the conversion since otherwise it was difficult to know if criticisms by Farley et al. (1998a) were legitimate. The use of spline conversion does seem to represent a complex handling of the original data, especially since similar conclusions had been reached earlier without their use (Lewis et al., 1996a).

Both the spline conversion and conclusions drawn from it have nevertheless been defended by Suissa and Spitzer (1998). Their stringent comments including a need for spline conversion because significant variation existed in distribution of use between controls in the first year of use, an essential need for a constraint to equalise risk of VTE in users and non-users at time zero use, and because the unconstrained model applied by Farley et al. (1998a) may ‘plausibly suggest’ that first-time users and non-users had different risks at time zero. They commented that the theory of the WHO authors ‘would indicate that in the WHO study, the baseline risk in first-time users of third generations pills was higher than that of second generation pills before they started to use their respective OCs. Such differences in risk at baseline could well explain why third generation pills were found to have a higher risk in that study’. They added that numbers in the WHO study were so few as to produce an unstable result. Lewis (1999) lists the advantages of spline conversions for first-time uninterrupted use and duration of use of OCs. These characteristics become effect modifiers rather than confounders in analyses carried out within strata. He comments that the WHO analysis of Farley et al. (1998a) included only 20 VTE cases and 12 controls distributed among several age groups who were first-time users of OCs, so it must have been less stable than the Transnational data with its comparable 39 cases and 62 controls. Lewis (1999) accepts that splines are ‘black boxes’, and that neither analysis was designed to cover first-time OC use, so studies should be specifically designed for this purpose. He comments in the present publication that ‘Given the situation at hand and the richness of the dataset, the ‘parsimonious interpretation of these data so eloquently advocated by Walker (1998)... which accepts a small increased risk of newer OCs, did not seem appropriate’. It is worth noting that the majority of criticisms of Transnational studies are more concerned with interpretations from its successive reinterpretations, rather than the criticisms of the original data base as aimed at WHO studies.

The term re-analysis is almost a perjorative word in epidemiological circles. Re-analyses are suspected of stretching data to reach further but not fully substantiated conclusions. Over-analyses can distort original data to produce unacceptable conclusions to those drawn initially, a criticism aimed at the Transnational study (Vandenbroucke et al., 1996). Re-analyses are fully acceptable if valid reasons exist, e.g. to apply new statistical methods to clarify existing data. This is why splines were applied to Transnational data (Suissa et al., 1997), so this re-analysis seems to be justified since adequate data were available and these conversions are used in other areas of epidemiology. Farley et al. (1998b) applied a new model to reassess data from earlier WHO studies and obtain time-related individualised estimates of VTE risk. This step also seems to be justified.

**GPRD Meditel and Mediplus studies**

Studies published in succession some time after the pill scare confirmed the trend to an equivalent or near-equivalent safety of second and third generation OCs for VTE risk as initiated in Transnational analyses. The GPRD analyses of Farmer (1996) and Farmer et al. (1997a, 1998), and the Danish study of Lidegaard (1997) point to weak risks of VTE among users of third generation OCs. The large GPRD database initially demonstrated significant switching of OCs, especially to third generation OCs, indicating how some users had significant problems with their previous OCs (Farmer et al., 1996). A note by Farmer (1996) initially utilised 37 cases of VTE, in a database of 347 257 women from Meditel over 12 months during 1990-1991, to report an equivalence of risk between second and third generation OCs. A larger and later MediPlus cohort analysis covered the years 1991-1995, with more cases and controls (Farmer et al., 1997a). A cohort analysis identified a slightly higher risk of ~1.3 for VTE with third versus second generation OCs (Farmer et al., 1997a), although the authors considered this analysis to be inadequately controlled for confounding by age. A related nested case control study with cases matched and controlled exactly by year of birth then identified no difference in risk between these different formulations, odds ratios being 0.8 and 0.9 for desogestrel and gestodene compared with levonorgestrel containing OCs. This nested analysis was among the first to match age by single years of birth from the outset of the study (Farmer et al., 1997a). It also adjusted for variables including history of current illnesses which remained unadjusted or incompletely adjusted in the cohort analysis, where some data used for adjustment, e.g. on BMI, were incomplete. Farmer et al. (1998) extended their MediPlus study to Germany, and again found no differences...
in VTE risk between second and third generation OCs. They point out the small size of this study (42 cases), the poor definition of VTE and absence of information on smoking and BMI although it was age-controlled by exact year and measured anticoagulant use as a diagnostic aid (Farmer et al. 1998). Farmer et al. (1997b) could not find any significant associations between market share and differences in mortality due to VTE over 11 years.

Extensive criticism has also accompanied these studies. Poulter et al. (1996, 1997) criticised these GPRD-based studies because the VTE cases were ill-defined and poorly diagnosed. They also questioned the very high VTE incidence, inconsistencies in the case control and nested analyses, how BMI was assessed and how co-variates were chosen for the log regression models in the study of Farmer et al. (1997a). Substantially lower risk estimates had been reported by Jick et al. (1995), so the higher estimates of Farmer et al. (1997a) could be a consequence of including diagnoses on outpatients, where errors are higher and many potential cases would have been excluded after a full clinical diagnosis. Under these circumstances it is essential that only idiopathic cases are included (Jick et al., 1998a). Farmer et al. (1999) responded that risks reported by Jick et al were lower than perhaps for any other published study, and the reasons for such low estimates should be investigated. Lidegaard and Milsom (1996) comment that the use of community controls could explain how the AAH Meditel study of Farmer (1996) identified smaller differences in VTE risk than other published studies.

Poulter et al. (1996, 1997) also stressed the lack of evidence on how matching by exact year of birth was achieved by Farmer et al. (1997a). They queried differences in the case control and nested analyses and why discrepancies arose with other reports. They pointed out how close matching by age had been tested previously, since most cases and controls fell within 2 years in WHO data, and within single years for some cases and controls in the study of Jick et al. (1995). Matching for exact year of birth in the WHO and Jick analyses was performed post hoc, so all controls differing by >1 year from their respective case in the cluster had to be excluded. Some controls could have been lost and some cases orphaned by this procedure. Adjusting ORs for this interval in the WHO study actually raised relative risk of third versus second generation OCs to 2.2. Poulter et al. (1997) pointed out how matching controls by 5 year bands combined with a sharply rising VTE incidence with age would be expected to overestimate prevalence of third generation OC use among controls and underestimate their adverse effects.

The GP computer-recorded diagnoses of VTE and anticoagulant therapy used by Farmer et al. (1997a, 1998) to define cases were also criticised as being uninterpretable by Jick et al. (1997). Vandenbroucke et al. (1997) and Jick et al. (1997) could not accept strong residual confounding by age as explaining either their 1995 data or WHO data (1995b), and Vandenbroucke et al. (1997) also criticised the log model of Farmer et al. (1997a) as being overadjusted, incomplete since data on BMI and familial occurrence should have been collected as prescriptions were given, and that unnecessary variables were included in the analysis. Matching by single year was done in their own Leiden study without exerting much effect on adjusted rate ratios. Farmer and Lawrensen (1997) did not contest that their diagnoses could sometimes be incorrect, but perhaps at no greater level than in other studies, and that their data on cases and controls was as reliable as in other studies including that of Jick et al. (1995). Anticoagulant therapy was a good measure since it demanded close attention and recording. Farmer et al. (1999) present a further analysis on single year matching in this symposium.

Recent epidemiological studies

We now reach the most recent investigations in our assessment of current data on safety of oral contraception. These have added to knowledge on the comparative safety for VTE and stroke of second and third generation OCs. The main thrust of our current paper is concerned with VTE, but it is necessary initially to discuss AMI briefly. Transnational studies had previously indicated a beneficial effect of third generation OCs on stroke (Lewis et al., 1997). The authors pointed to the large size of their study as one factor in this discovery. Lidegaard and Kreiner (1998) recently reported on risks of cerebral thrombosis and OC use from a Danish database. ORs for users of first, second and third generation OCs were 1.9 (0.9–3.9), 2.4 (1.4–4.2) and 1.3 (0.8–2.2), leading the authors to comment on lower OR with third generation OCs versus second generation OCs, and to oestrogen concentrations persisting as significant risk factors. Dunn et al. (1999) found no differences between OC type and risk of MI in the UK [ORs of 1.10 (0.52–2.30) and 1.96 (0.87–4.39) for second and third generation OCs respectively]. Their values agreed with Transnational U.K. data, but not with the greater safety of third generation OCs in the Transnational European study (Lewis et al., 1997). Two further studies amend the Transnational findings by reporting no difference in risk with second and third generation OCs (Jick et al., 1999; Poulter et al., 1999). Further clarification is needed on this point, although these differences between studies do not challenge overall conclusions by most or all epidemiologists that risks of stroke are overall similar with both OC generations.

VTE continues to be the main source of disagreements over the relative safety of second and third generation OCs, and will be the major focus of concern in the rest of the present paper. The concept of a higher risk with third generation OCs above levonorgestrel OCs is re-stated in two recent WHO-based publications. The WHO Technical Report (1998) comments how current users have a low absolute risk of VTE, but this is nevertheless 3–6 times higher than in non
users, and probably highest during the first year of use. Third generation OCs containing desogestrel or gestodene have a small extra risk above other OCs. Contributors to a WHO meeting, published in *Contraception*, continued the theme of the higher risk with third generation OCs (reviewed in Edwards and Cohen, 1999). Among these papers, a fresh review of international data and a new model were applied to assess overall risks of second and third generation OCs at various ages in individual users, based on the original WHO data (WHO 1995a, b). The doubled risk of third generation versus second generation OCs was confirmed, and the model proved valuable in utilising overall data for all OC-related illnesses to calculate their combined risk to users at all ages (Farley et al., 1998a, b). Outcomes for VTE seem to be virtually unchanged from earlier WHO analyses, and no corrections have apparently been made to the original data base for the bias or confounding identified in other large scale studies. WHO-based discussions thus continue to stress the robustness of their data and argue that biases and confounders are too weak to modify the conclusions drawn in 1995 (Walker, 1998a).

WHO contributors to the present symposium largely reiterate their earlier beliefs on OC safety by questioning the significance of various biases in recent studies which came to the opposite conclusion (Farley et al., 1999). They point out how prescription bias, i.e. the preferential prescribing of certain OCs to women at risk of the disease under study, can arise under specific conditions and be assessed indirectly by measuring associated variables. The consistency of prescription patterns was indeed lower in some studies than initially claimed by physicians (Heinemann et al., 1996). This factor was adequately covered anyway in some early case control studies by adjusting for age. Selective prescribing was virtually absent from WHO studies and perhaps of minor importance in relation to recently-identified genetic risk factors for VTE. The design of modern studies could cover these recently discovered risks. Farley et al. (1999) also stress how diagnostic bias, notoriously difficult to control, could arise if users of particular OCs were referred for additional tests of risk factors. Its underlying causes must be independent of characteristics recorded in the study, and it was not very significant when certainty of diagnosis was measured. Attrition of susceptibles could be related to higher risks in first-time or younger users or to OC switching. Farley et al. (1999) question any empirical support for large-scale losses of susceptible groups, and comment that age matching might largely control for it anyway. Its rarity also seems inadequate to explain the higher ORs associated with third generation OCs. They also doubted whether age matching by exact year of age or patterns of use reduce risk ratios as claimed by Farmer et al. (1997). Exact matching by year (Jick et al., 1997; Lidegaard et al (1998b) or 2 years (Poulter et al., 1997) made little difference to estimates gained from wider age bands. The WHO team also glanced back to the spline controversy in questioning the significance of duration of OC use (Farlet et al., 1999). They reiterated the very low risks of OC-related stroke especially after blood pressure checks, confirm how OCs account for less than half the low levels of total cardiovascular risk, and stress the very high overall absolute safety of all recent OC formulations.

Some of these comments of the WHO group are sustained in recent studies by Herings et al. (1999) and Bloemenkamp et al. (1999). Herings et al. (1999) conducted a small scale Dutch study using computerised data on pharmacy records and hospital admission among 450,000 residents. This was linked to nationwide hospital discharge records of high specificity and sensitivity, to provide a base for searching for data on OCs between 1986 and 1995. New OC use was assessed among women aged 15–49 who ever used OCs between these dates and without prior evidence of VTE, use of anticoagulants, depot hormones, cardiovascular drugs, morning after pills or hospitalization 2 months before starting OCs. First episodes of exclusive use of levonorgestrel, gestodene and desogestrel OCs were identified, and hospital discharge records searched for first occurrence of VTE. Logistic regression models revealed risk ratios of 9.0 and 2.4/10,000 person-years with third and second generation OCs respectively (relative risk 3.5; 1.4–8.8). Highest relative risks occurred in healthy young women aged <25 with no evidence of pregnancy. Highest absolute risks arose in the first year of use of third and second generation OCs (32.1 and 12.4/10,000 person years respectively), and the relative adjusted risk ratio for VTE was 14.1 (1.9–106.4) during the first year after exclusion of any chronic diseases, excess weight and psychiatric distress. Adjusted relative risk declined to 1.7 (0.9–3.1) in longer-term users. Matters of concern in this study include risk ratios which are far higher than in any previous large-scale study, very wide confidence intervals which indicated no difference between second and third generation OCs in later years of use, and low case numbers of cases of 27 with third generation OCs and six with second generation OCs. Clearly well aware of the possibility, Herings et al. (1999) wondered if keen marketing of third generation OCs had encouraged their use by highly susceptible users, or if unidentified prothrombic mutations influenced their results.

The study by Bloemenkamp et al. (1999) was performed in two clinics on referrals for suspected DVT of the leg. Totals of 198 patients and 591 controls with a first VTE episode were assessed, both groups being subjected to the same referral and diagnostic procedures. Overall ORs for OC use was 3.2 (2.3–4.5), rising to 3.9 (2.6–5.7) when adjusted for age, family VTE history, age and centre. Idiopathic patients had very similar rising ORs of 3.8 (2.5–5.9) for OC use and 5.0 (3.1–8.2) after adjustment. Its authors conclude that diagnostic suspicion and referral bias do not have an important role in epidemiological studies on safety of modern OCs.
In complete contrast to studies proposing higher risks of VTE with third compared with second generation OCs, several new studies reported in 1998/1999 conclude that each of these OCs carry similar risks for VTE. Lidegaard et al. (1998a,b) reported a case-control designed study to estimate the duration of use of OCs and other factors involved in estimating OC safety, following earlier work of Lewis et al. (1996a). In it, cases and randomly-selected controls were identified from all Danish women aged 15–44 years with DVT or PE. Potential confounders included BMI, length of use, family history of VTE, AMI or stroke, smoking, coagulopathies, diabetes, years of schooling, certainty of diagnosis, previous births and (incompletely) hypertension in pregnancy. Details of the graphical log linear model for multidimensional contingency tables and validity of the analysis are given in the original paper. Odds ratios declined sharply with time: 5.1 for <1 year, 2.5 for 1–5 years, and 2.1 for >5 years, as compared with non-current users. Adjustments for OC type did not alter the significance of this trend for duration of use. Second and third generation OC users had ORs of 1.8 (1.1–2.9) and 3.2 (2.3–4.4) respectively for VTE which became 0.9 (0.4–2.0) and 1.3 (0.6–2.7) when adjusted for duration of use, so indicating a healthy user effect. Many short- and long-term users were taking third and second generation OCs respectively. An indication of prescriber bias also emerged from the data. The authors believed they had identified a primary cause of higher VTE risks reported for third generation OCs, in agreement with earlier conclusions from Transnational and Danish databases, and had shown that third generation OCs had similar risks as second generation OCs (Lidegaard et al., 1998b).

Even though published so recently, the study was soon critcised as Walker (1998a) commented that data on risks were too closely linked to third-generation OC use. Lidegaard et al. (1998c) defended the wide basis on large sections of the Danish population, its careful checks for bias and far greater precision than in the WHO study of Poulter et al. (1996). Walker (1998b) again disagreed, and the WHO Scientific Group (1998) and Farley et al. (1999) concurred with him because short-term risk was concentrated in third generation users, very few women used levonorgestrel OCs over short periods and data on cases and controls were collected at different times for cases and controls. Lidegaard et al. (1998b) had selected controls for studies on stroke, which occurs at older ages than VTE, and applied them to their VTE study. Age adjustment therefore became essential. Moreover, Farley et al. (1999) stressed how patterns of use and type of OC differed according to age, as typified by the greater use of third generation products for <1 year as compared with levonorgestrel OCs. Effects of duration of use would thereby be confused from those due to OC type.

Lewis et al. (1999) and Lewis (1999) in the present symposium report a very recent analysis of Transnational data on VTE risk including new data from users of second and third generation OCs. An enlarged database provides more information on healthy user effect for the VTE risk previously identified using standard logistic regressions and spline conversions for duration of use and past OC exposure. Recorded data included all past exposures among 502 cases and 1864 controls, full exposure history by month since menarche, type of OC, pregnancies and periods of non-exposure analysed from 17 622 continuous exposure periods covering 47 914 person-years of observations (Lewis et al., 1999). Risk analyses were applied via a logistic regression model with time-dependent covariates, itself a novel application to safety of OCs. The conditional probability that a failure occurred in a case rather than in a different member of the risk set was computed. Odds ratios of 0.79 (0.50–1.26) were obtained for third versus second generation OC users. Recall bias was unlikely to have influenced this calculation as shown by sensitivity checks. This new evidence stresses again how bias, in this case several aspects of healthy user effect, explains why initial Transnational estimates indicated higher risk ratios with third generation OCs. Its authors conclude that no differences exist for VTE risk or any other outcome among users of second and third generation OCs, and their data strongly confirms their earlier and similar conclusions. They stressed how case control studies usually assess well-defined events occurring soon after drug use whereas VTE events are ill-defined and arise after previous high risk exposures. Their approach included past responses and precise timing of duration.

In the present symposium, Farley et al. (1999) question such recent ‘improved’ adjustments of epidemiological data including that by Lewis et al. (1999). They also query aspects of this extended Transnational study with its inclusion of extra data to completed OC use, and request full details of the model. Risk can be enhanced among future cases exposed to relevant risk factors, so the relative risks between OCs could be attenuated to 1 especially with high relative risks. Moreover, Lewis et al. (1999) identified no differences of prevalence of past use of different OC formulations between cases and controls, including non-users, which questions how past exposure explains excess risk with third generation OCs. Lewis (1999) in the present symposium and Lewis et al. (1999) discuss the use of the Cox proportional hazards time model, some limitations in its application and how it conforms with similar trends gained by standard logistic regressions.

A third recently-available contribution to debates on safety of third generation OCs, was based on an extension of the Mediplus study. This more stringent nested case control analysis added new data collected until mid-1997 on ‘true’ contraceptive history for each woman (Todd et al., 1999). Cases were diagnosed for DVT or PE, anticoagulation treatment, and exposure to OCs on the day of the VTE, and were verified. Controls were case matched by year of birth and GP practice. A total of 99 cases were available for analysis, including 61 from the original study. Crude odds
ratio for idiopathic VTE among users of all OCs was 4.6. Relative values calculated against low-oestrogen levonorgestrel OCs as a reference revealed similar odds ratios of 1.1 (0.5-2.4) for desogestrel and gestodene OCs. This set of investigators conclude that third generation OCs hold no extra risk for VTE above those typical of second generation OCs. Farley et al. (1999) question some aspects of this study of Todd et al. (1999), which spanned a longer time frame and had more cases than their previous investigation. Adjusting for number of non-OC prescriptions 6 months prior to index date might influence conclusions drawn from the study if prescribed for conditions typical of an early manifestation of the disease in question.

A fourth study involving a vigorous defence of an earlier use of exact year of birth in assessing OC risks for individual formulations is briefly summarised in this symposium by Farmer et al. (1999). Two large cohort/case control studies based on Medilplus and GPRD databases involved comparisons with data of Todd et al. (1999). Carried out during a period of increased OC switching from third to second generation OCs following the U.K. Committee's announcement, its crude incidence rates indicated lower odds ratios for levonorgestrel OCs than for desogestrel or gestodene OCs, but not after matching controls to GP practice and year of birth. Consistently higher ORs for third generation OCs arose with matching by 5-year age bands. Farmer et al. (1999) comment how significant ORs for higher VTE risk with third generation OCs emerged in earlier studies, e.g. WHO (1995), after combining data for gestodene and desogestrel OCs. Even then, they were significant only for hospital and not community controls. Likewise, Transnational U.K. data identified no significant differences between levonorgestrel versus gestodene OCs, and a barely significant OR of 2.6 (1.7-2.2) in German data. Comparisons for desogestrel were non-significant in Germany and marginally elevated in U.K. (Farmer et al., 1999). Highly significant ORs emerged when data were aggregated from both third generation OCs. The authors conclude that their studies do not indicate higher risks of VTE with any single recent OC formulation.

Deciding on the relative safety of different OC formulations

This brief review of epidemiological studies and controversies from 1995 shows how conclusions drawn from the four original studies showing a higher risk of VTE with third generation OCs have been modified by extensions to the Transnational study and further work on Meditel, Medilplus and GPRD databases. A quartet of new data in 1998/9 continues the trend and stress the equivalent safety of third generation OCs. Comments and criticisms on these papers will no doubt emerge, since as Westhoff (1996) comments, each new epidemiological study leads to controversy as its strengths and weaknesses are probed. This epidemiologist suspects that numerous analyses and re-analyses of large amounts of data may never formally solve the importance of bias and confounding in complicated risk estimates. Clarification will not be easy since absolute risks of some diseases are now so small that enormous trials are necessary to provide meaningful data. Lewis and Spitzer (1997) calculate that a doubling of relative risk of VTE will require choosing and assessing 300 000 women for 1 year, while a four-fold increase in relative risk will require trials involving 60 000 women. Heinemann (1999) highlights this situation in the present publication. The rarity of cases alone implies that absolute risks are very low, and probably far less than many ordinary duties in home or office.

The lack of a uniform opinion after so much work raises questions about epidemiological methods and interpretations, and if they are sufficiently reliable to dispel the highly critical remarks written about them in Science (Taubes, 1995). No gold standard exists to measure and compare the validity of successive studies on oral contraceptives. Can a compromise on OC safety be reached, e.g. by calculating weighted mean of all studies, complete with SE and based on variations between the best studies? An underlying homogeneity among selected studies would be essential for this purpose, yet untested variables such as the genetics of different populations, methods of diagnosis especially for VTE, and variations in medical practice in different countries could introduce variations between them. This approach is highly unlikely to be of any help for other reasons. Constructing some sort of mean implies making judgements on the value of individual studies, or applying weighting methods according to size or other factors, and which studies to include in the mean. This is apparent in the calculations of Farley et al. (1999) in this publication, where weighted average rates of VTE risks are derived using data from eight studies published between 1995 and 1999. This produces an overall weighted average of 1.9 (1.5-2.2). Unfortunately, the authors chose to include only the original 1996 data from the Transnational study, when several new analyses derived from Transnational data would have reduced the overall average. They also included three 1999 studies, two of which were relatively small. One of them included two cases exposed to second generation OCs and six exposed to third generation OCs, with an RR of 1.8 (0.3-11) (Vasilakis et al., 1999). Exposed cases in the other small study numbered six with third generation and 27 with second generation OCs, producing an RR of 4.2 (1.7-10), which is far above any of the other studies listed.

If any sort of value judgement has to be made, it will in our opinion involve a shift in attitudes toward similar relative risks with second and third generation OCs, from the moderately higher relative risks with third generation OCs reported in the four 1995/1996 studies. The tabulated data of Farley et al. (1999) displays a gradual declining trend in RRs from 1995 to 1999, except for one of the small 1999 studies (Herings et al.,
specifically aimed at assessing relative risks of second and third generation OCs. Except for Herings et al. (1999), none of the new studies confirms the twofold difference in risk typical of the 1995/1996. Yet even then their reported extremely high differences between generations of OCs in the first year are succeeded by no such differences in later years. Other aspects of this study require close examination including some extremely wide risk ratios.

The overall balance of new data has shifted to an equivalence of safety for second and third generation OCs. It is doubtful if subsequent criticism and argument over recent equivalent risk rates will change this conclusion. Perhaps the lowest common denominator of agreement, shared by many of the world’s premier investigators we have interviewed, is that absolute risks of all OCs are very low, and differ hardly at all between various formulations. All statements on relative risks of different formulations qualify the earlier higher risks reported with third generation OCs. This is seen in recent statements from several highly professional groups. The ESHRE Capri Workshop (1998) gathered many distinguished epidemiologists, several of them authors on previous WHO, Transnational and other relevant studies. This group identified no significant increased risk of MI with OCs or hormone replacement therapy (HRT), or between cerebrovascular disease and OCs. In relation to VTE, they recommended women at risk not to use OCs, a measure already widely practised, and that women not at increased risk should not be deterred from OC use or from continuing with a satisfactory formulation. They also accepted that there was insufficient evidence of biological mechanisms which could enhance VTE risk. We note the UK Medicines Commission, an independent body, has apparently modified the original opinion on increased risk and lifted some restrictions on prescribing third generation OCs imposed in 1995 (Metters, 1999). This decision was subsequent to an appeal by manufacturers, and opposed by the Medicines Control Agency. The American College of Obstetrics and Gynecology (1998) reconfirmed higher VTE risks arose with desogestrel OCs used in the USA, but defers to decisions of individual clinicians and patients who may benefit from some formulations (Weiss, 1999). The French Agence de Medicament accept that any differences between OC formulations are so slight that contraindications should not differ between them. The Administrative Court of Berlin (1997) considered that third generation OCs were no less safe than second generation OCs.

We accept that conclusions on VTE risks in the original studies have been modified as several other large scale trials have searched for bias and confounding. To do otherwise, or to dissent from our opinion, would involve questioning the great majority of studies reported since 1995, each specifically aimed at assessing relative risks of second and third generation OCs. Investigators of the largest of these studies, the Transnational, have modified their own conclusions from that time based on successive analyses initially included in the study protocol. Spitzer (1998, 1999), the senior investigator of this study, summarises the opinions we have drawn in this present manuscript. He believes the initial studies were affected by three biases: prescription, referral and healthy user, which have been corrected in later analyses. He stresses how, among all these databases, third generation OCs were usually prescribed to users who were short-term, younger, with personal or family histories of VTE, clinical venous signs, chronic inflammatory disease and seven other risk factors. In passing, it is worth noting the personal opinions of many gynaecologists agree with these remarks as they recall placing highly obese women and others with risk factors on third generation OCs, although such received opinions are not always confirmed in subsequent epidemiological analyses (Heinemann et al., 1996; Farley et al., 1999). Spitzer (1999) comments on how the WHO Technical Report (1998) failed to sufficiently recognize the corrections for bias and confounding in later reports, and criticises it for what it omitted to say: that even if a small extra risk did exist with third generation OCs then absolute risks of various sequelae including death ‘are infinitesimal, of no clinical importance and definitely of no public health significance’. No evidence has emerged of any epidemic of a cardiovascular disease attributable to an OC, he asserts, and emphasises that OCs containing desogestrel or gestodene do not carry even a small additional risk when compared with second generation OCs. He re-stresses the very low absolute risk identified in the four original studies, and contrasts this fact with its alarmist presentation (‘doubling of risk’) to the public in 1995 by regulatory authorities and some scientists.

We note too that Lidegaard et al. (1999) accept that bias and confounding explained most, but possible not all of the reported differences between second and third generation OCs in the original studies. They also indicate that young women without risk factors can use any low-dose OC with equal safety, although existing risk factors for arterial disease might preferably indicate the preferential use of third generation OCs whereas factors for venous disease might imply a preference for second generation OCs. For patients aged >30 years, third generation OCs confer fewer thrombotic effects.

We also agree with Farley et al. (1999) when they call for new studies on interactions between hormones and haemostasis, so that even safer OCs can be produced. The need for research in this field has not lessened, for example on the curious higher VTE risk with lower than with higher oestrogen formulations, or on the need for clear biochemical data on the mode of action of all OCs. This point is clearly made by Heinemann (1999) as he searches for new horizons to clarify special aspects of pharmacoepidemiology of oral contraception. Indeed, dosage may be a significant factor,
since the lower levonorgestrel levels may explain why multiphasic levonorgestrol OCs invoke a higher resistance to activated protein kinase C among users than their monophasic counterparts (Kluft et al., 1999). Concentration-dependency may explain some effects of all OCs, and will have to be taken into account in subsequent epidemiological studies on all second and third generation formulations. In this field, risk estimates are greatly influenced by knowledge, behaviour or intentions of users and investigators. They are also influenced by mild symptoms and misclassifications which may result in 50% of VTE cases being missed. Heinemann introduces a need for 'unmasking bias' as a further work-up of patients and controls to avoid systematic errors. He stresses how genetic predisposition or family histories for VTE must be assessed, previous thromboses not treated in hospital and OC switching should be clarified, excluded cases given more emphasis, length of OC availability be recorded, and blood samples taken more regularly because fundamental information has been missed in the past.

We conclude that sufficient evidence now exists to demonstrate an equivalent safety of third and second generation OCs. We list our conclusions and recommendation below. Before then, we comment briefly on how the pill scare arose and how further scares may be avoided.

How the pill crisis arose

This latest pill scare is the last in a long series. Will it be the last, or is another just around the corner? For present purposes, the recent furore is still with us, in the minds of many doctors, epidemiologists and women. The epidemiological evidence assessed above indicates that sufficient evidence did emerge suddenly in 1995, and enough to justify some caution about the prescription of third generation OCs. Was the resulting turbulence in scientific, clinical and social circles justified, and could it have been avoided? Sufficient comment has been made on the social effects of the pill scare (Szarewski and Mansour, 1999; Mills and Edwards, 1999). There were many players other than those involved in the social upsets, and we wish to restrict our comments on the origin of the scare to regulatory authorities, pharmaceutical companies, and then to the numerous other organisations including health authorities, Press and administrators.

The wider social background of the pill scare

The panoply of events which initiated, amplified and, we trust, now mitigate the pill scare include comments or actions of professionals in all walks of life. The numerous interactions between them and their immediate consequences can be traced only briefly in this article. The events are international in scope and make a fascinating story of interactions between modern epidemiological research, responses of regulatory authorities and reactions among very wide sections of society. In sequence, the events occurred as follows.

In 1986, WHO undertook epidemiological studies to judge the current safety of OCs worldwide because this information was lacking especially in developing countries. Their study was not initially designed to compare VTE complications with second and third generation OCs. Their initial results seemed to indicate a raise in risk of VTE for third generation compared to second generation formulations in a small subset of the data, mainly from Oxford.

Without notifying some contributors to the studies, and after data were submitted to journals but before its publication, analyses of WHO data became available to the UK Committee of Safety of Medicines. The CSM had joined in commissioning the Jick study and possessed initial results of the Transnational study. After 6 h of non-published discussion, this Committee decided to issue a circular to doctors. The Committee was probably under considerable public and possibly even legal pressure. It might have feared that failing to make an announcement warning against an impending catastrophe could well have been worse than the pill scare itself if the catastrophe had come true. With hindsight, it now appears they should have limited themselves to warnings rather than to an action virtually amounting to prohibition of some OC formulations.

This information was prematurely broadcast on BBC and published in the Press. It led to news bulletins such as a declaration that thromboembolic complications arose with new low dose combined oral contraceptives. According to such reports, authorities regulating public health possessed details of twenty documented cases of thromboembolic complications in users of third generation OC users in the UK and Federal Republic of Germany.

News spread worldwide. It appeared to cause a pill panic especially in UK, Germany and Norway. Many users in these countries abruptly terminated the use of OCs, resulting in a later increase in numbers of abortions. Even today, a certain number of sexually active young women have not resumed OC use. In other countries, in Europe in particular, health authorities were more cautious and decided to withhold any decision until the data were ready for examination. This moderating attitude did not prevent some patients from being worried about OC use. Finally, an international debate began. Biases appear to have distorted some of the conclusions reported in the various analyses. We have reviewed these events in this article and in our Introduction, and believe they now indicate no additional risk of VTE in users of third generation OCs. Unanimity has not yet been reached by all epidemiologists. Most of them agree that OCs are very safe especially in absolute terms, and that women should use them without notifying some contributors to the studies, and after data were submitted to journals but before its publication, analyses of WHO data became available to the UK Committee of Safety of Medicines. The CSM had joined in commissioning the Jick study and possessed initial results of the Transnational study. After 6 h of non-published discussion, this Committee decided to issue a circular to doctors. The Committee was probably under considerable public and possibly even legal pressure. It might have feared that failing to make an announcement warning against an impending catastrophe could well have been worse than the pill scare itself if the catastrophe had come true. With hindsight, it now appears they should have limited themselves to warnings rather than to an action virtually amounting to prohibition of some OC formulations.

This information was prematurely broadcast on BBC and published in the Press. It led to news bulletins such as a declaration that thromboembolic complications arose with new low dose combined oral contraceptives. According to such reports, authorities regulating public health possessed details of twenty documented cases of thromboembolic complications in users of third generation OC users in the UK and Federal Republic of Germany.

News spread worldwide. It appeared to cause a pill panic especially in UK, Germany and Norway. Many users in these countries abruptly terminated the use of OCs, resulting in a later increase in numbers of abortions. Even today, a certain number of sexually active young women have not resumed OC use. In other countries, in Europe in particular, health authorities were more cautious and decided to withhold any decision until the data were ready for examination. This moderating attitude did not prevent some patients from being worried about OC use. Finally, an international debate began. Biases appear to have distorted some of the conclusions reported in the various analyses. We have reviewed these events in this article and in our Introduction, and believe they now indicate no additional risk of VTE in users of third generation OCs. Unanimity has not yet been reached by all epidemiologists. Most of them agree that OCs are very safe especially in absolute terms, and that women should use them consistently and not to be afraid of using any of them if contraindications have been excluded.

Several other factors could have played a role in the pill scare, and explain why these successive events arose. One of
them could well be the link between OCs and sexuality, and the dangers of once again raising objections from those hostile to a liberated sex life in some countries.

Role of regulatory committees on the safety of oral contraception

In relation to regulatory authorities involved in the pill scare, we believe that the UK and German Committees made a serious error. We could well have agreed with their assessment of the available data on VTE, since they based their actions in October and November 1995 on data from three or four studies, i.e. WHO, Jick and Transnational and the Leiden study became available in December 1995. These epidemiological teams indicated similar estimates of higher relative risk with third generation OCs. Things might have been very different for the Committee if even one report had reached a different conclusion. This was not to be since the 'black week' for OCs, as Heinemann (1999) calls it, was heralded by four nearly-simultaneous studies carrying the same message.

We do not know if the U.K. Committee was being pressurised to make their decision and relay their conclusions so widely. Even today, uncertainty still surrounds OC safety, yet they acted immediately even though drugs now have 'safety profiles' rather than 'risk factors'. Reservations over its decisions among some original investigators of these databases did not reach the Committee or were overlooked. Outside factors might have interfered with its decision, such as the higher costs of third generation versus second generation OCs on limited budgets. Costs are not reimbursed for third generation OCs by insurance organizations in many countries including France because of their high price. Even practitioners and others in favour of third generation OCs had reservations about such a great difference in prices of second and third generation pills which varied from country to country. Nevertheless, facts such as the limited degree of risk even on the most critical of the four original analyses, the preliminary nature of some of them, and the possibility of confounding and bias should have led the UK Committee to think twice before making such a critical decision, let alone issuing it. Nor can we condone the response of the German Committee, which has also been roundly criticised and penalised by the German judiciary. It is a sad day when regulations on the safety of OCs have to be judged by due process of law rather than in responsible scientific and clinical circles. Alternative forms of action were available which would could have avoided the scare (Edwards and Cohen, 1999), and the more gradual steps open to them would have been preferable to the one they chose. A rapid implementation of their decision compounded the original error.

Pharmaceutical companies

We have no evidence of any actions by pharmaceutical companies which may have led to the pill crisis. They have an immense financial stake in the sale of OCs, perhaps the most profitable drugs on the market as in the USA where contraceptive sales are estimated to be worth US$2600 million. Nor is there any evidence of any role for the pharmaceutical companies in influencing analyses on safety of OCs. Schering funded the Transnational analysis when a scare arose in Germany, at the specific request of German authorities. Organon approached us to prepare this present publication, but without any payments or compromising our complete editorial freedom.

Company reports and databases are often viewed with doubts about their veracity and suspicions of special pleading. Yet companies must make and respect objective assessments of trial data on their products. Their best interests are clearly served by frankly recognising the advantages and disadvantages of their products because any problems will soon be identified by outside agencies, especially over a topic such as the safety of such a huge number of OC users. No company could afford to hide any dangers of OCs, and then face the music when those dangers were revealed openly. Indeed the only question needing clarification concerning the commercial distribution of OCs is why third generation products became associated in the minds of so many doctors with greater benefits for users. Did a too-aggressive marketing ploy, on issues such as a 'metabolically neutral profile', lead to this state of affairs? Lidegaard and Milsom (1996) believed third generation products were marketed as being safer than second generation, and indeed being so safe that users with contraindications were prescribed. Apart from this possibility, we believe that the companies have acted even with fortitude at a time when they were losing thousands of customers during the pill scare. Perhaps their losses will never be recouped, especially those companies who were producing third generation OCs only.

Can other scares be avoided?

Avoiding pill scares may never be easy. The safety of OCs is an essential aspect of clinical care and its regulation, of interest to most countries worldwide. Experience of successive scares seems to indicate that others are likely, for whatever reason. Regulation operates effectively and smoothly for most of the time yet the odd hiccup spreads widespread concern and jeopardises wellbeing among users. Specific aspects of current regulatory systems seem require attention and revision, and some steps can be taken to achieve this aim.

First, the familiar interacting sequence of events during OC scares reveals how, once started, more and more commentators, critics and Press rapidly amplify the details.
One essential point of control is at the level of decision making about a new safety risk. Closer control there could avoid or modify the immediately following events. Possibly announcements on OC safety should not be made until a cooling-off period has elapsed and other authorities have been consulted. Manufacturers should be consulted about the safety profiles of their products. Other epidemiologists – especially those donating their data – should have been consulted before the past pill scare erupted. Essentially, extreme care is expected in dealing with such a sensitive topic and making public pronouncements. In the widest interests of society, it would be preferable to ensure that regulators and committees have an easy access to the widest advice available, and care is taken to ensure how their decisions are examined in detail and quickly by peers chosen from epidemiological and social scientists before being issued to the public.

Second, it is essential that sufficient guidance or adequate controls are placed on regulators. Some form of discipline may be needed, which might become counterproductive if applied too harshly or if it hampered the making of open and free decisions. Do they have clear guidelines available which could help to ameliorate taking or acting on an unwise decision? Perhaps very well-defined yellow, orange and red alerts, each described in detailed clinical, epidemiological and social terms must be made available and compulsory. Some forms of warning of this kind are already practised in some countries. For example, when there is a doubt about a product in France, the Commission de Pharmacovigilance makes an enquiry and issues a report. This was the case for a recent drug named Atymil when the Commission decided there was no reason to change the actual type of prescription. Perhaps very well-defined yellow, orange and red alerts, each described in detailed clinical, epidemiological and social terms must be made available and compulsory. Some forms of warning of this kind are already practised in some countries. For example, when there is a doubt about a product in France, the Commission de Pharmacovigilance makes an enquiry and issues a report. This was the case for a recent drug named Atymil when the Commission decided there was no reason to change the actual type of prescription. If there is a strong convergence of dangerous secondary effects, as for anorexigens, the Commission may take temporary measures and issue a provisional suspension of the drug or seek information about its secondary effects. If there is a very severe risk, the Commission issues a red envelope to all pharmacists who must remove the drug from circulation in an emergency procedure. Perhaps procedures such as this might have helped the U.K. Committee who might then have issued warnings without taking action.

Third, regulators include civil servants, members of commissions, health organisations, and independent agencies, each with the power to make comments or take decisions with potentially disastrous effects. All these agencies are actually making clinical decisions, yet unlike practitioners many of these regulators cannot be prosecuted for making decisions which are disastrous, e.g. for users of OCs. They may have to obey a written law or emerging guidelines, but do not consult with a patient, face a General Medical Committee, or risk being charged in a court of law for malpractice.

Fourth, Press Officers are essential and must be knowledgeable. They should be a normal content of committees, clearly non-governmental, and have authority to contact other regulatory authorities in advance of any critical statement being issued. An easily-accessible source of information and advice could help to dispel some of the wilder rumours and actions usually associated with a pill scare. They would have helped to avoid the whole pattern of events of 1995 pill scare which revealed once again a weakness in the manner whereby fundamental fears for public safety are handled and communicated (Furedi, 1999). This is not the first scare on the pill, or indeed on other epidemiological factors affecting the lives of many people. Yet it brought fundamental fears of many users into sharp perspective since insufficient care was taken to handle the matter sensitively.

Fifth, we have no doubt that the Committees in the present pill scare acted in good faith, but their actions caused immense distress. We regret their refusal to give their point of view to this symposium since any help from them could have been invaluable. Did they act wisely in issuing their ‘Dear Doctor’ letter? It was to lead to distortions as its news was transmitted in various ways and received by an unprepared Press and by widely varying organisations, each with a stake in the safety of oral contraception (Mills and Edwards, 1999). The fact that regulatory authorities in many other countries did not follow the U.K. lead surely shows that wider consultation is essential. In fact, these distant authorities faced a potentially larger problem than the British since their decision was taken after the scare had erupted and among widespread public concern.

Sixth, actions subsequent to the initial scare were not helpful. There was a feeling of disagreement among epidemiologists, clinicians and regulators. Clear advice was hard to get, and many OC practitioners were uncertain what advice to give to women. A clear central direction was essential, yet was lacking, and steps are needed to place such an authority in position. Even today, warnings are being proposed or placed on some OC third generation OC products at the discretion of some regulatory agencies. This is an unwanted step, adding more confusion to a situation already fraught with enough difficulties and contraindications. It is a different matter to placing health warnings on cigarettes where few epidemiologists question the dangers. Labels on some OCs will almost certainly add confusion to users, anger epidemiologists and others who consider second and third generation OCs to be equally safe, be contrary to widespread agreement that absolute risks are minute with all OCs, and penalise products that many senior epidemiologists consider to have no extra dangers.

Overall, there should be no surprise at the importance given to the four epidemiological studies concerning VTE incidents when they were published. Since the pill has been on the market, successive waves of panic have concerned cancer in general, skin diseases, cardiac incidents, breast cancer..... Everything then settles down a few months later as second thoughts prevail and all returns to normal. Publicity in the media can preclude the establishment or continuation of fundamental epidemiological studies, as in the case of the
Transnational study which had to be abandoned in some countries because of fears of media confounding.

It is far from easy for physicians to help their patients make rational decisions for their patients about such complex data in view of incessant debates and arguments about significance. When faced by anxious patients and uncertain advice from regulatory authorities, should they stop prescribing OCs, or continue to do so and so assume an immense responsibility and legal risk? Epidemiology becomes very distant, and physicians have to remember at such critical junctures that many forms of bias and confounding are operative in epidemiological studies. They risk prosecution for taking decisions that are questioned later, e.g. about continuing transfusion services during the onset of the HIV epidemic when the role of viruses was uncertain. Each successive OC scare weakens clinical authority and that of epidemiologists alike.

Conclusion: the essential background to the pill scare

The risks of OCs, including VTE and MI, have been known since the 1960s and are far less than those observed during pregnancy. Most risks are not very important clinically in the widest terms. Death from VTE, for example, accounts for <1% of total mortality in women aged 15–44 years and is predominantly associated with trauma, surgery and major illness.

The differences observed in the studies between second and third generation OCs are almost certainly explained by bias rather than by causal relationship. There is a known predisposition to VTE which can be either an inherited abnormality (100 different single gene mutations have been identified: Bertina and Rosendaal, 1998; Laffan and Tuddenham, 1998) or might be an acquired hypercoagulable state resulting from diseases, pregnancy or drugs. Any relationship with the use of particular OC formulations is highly speculative at the present time. The great majority of current evidence and opinion indicates that third generation progestagens have similar effects on those variables leading to haemostasis as with earlier low-dose combinations, differences between them being non-existent (Speroff et al. 1993; Stubblefield, 1993; Robinson, 1994; Winkler, 1999; Speroff, 1999).

Studies and debates on safety of third generation OCs have been carried out by the world’s most distinguished epidemiologists. Disputes in the literature coming from such authorities must be respected as revealing genuine differences of opinion. Such disagreements obscure large areas of agreement, such as considering any comparative study as valid only if women with a predisposition to VTE are excluded from control and user groups when estimating relative risk.

The epidemiological studies published in 1995–96 reporting a difference in the risk of VTE in women using 3rd generation OCs are highly speculative at the present time. The usual precautions in selecting appropriate candidates for OCs need to be applied. Women who smoke, who are obese, hypertensive, diabetic, or who have a personal or family history of thrombosis will need to have an individualised assessment of their risks of cardiovascular disease, whether or not they use oral contraceptives; only after this assessment can they make informed contraceptive decisions and modify their risk factors wherever possible. At this point in time specialised testing for coagulation disorders is discouraged. We also agree with the findings of the ESHRE Capri Workshop Group (1998), who wrote among other conclusions that: 'For women not at increased VTE risk, the likelihood of VTE should not be a deterrent to the use of OC, nor a reason to change from an otherwise satisfactory product'.

It is essential to relate odds ratios to crude absolute risks. Numbers, rather than percentages, are better understood by
doctors and public. With present state of knowledge, there is no reason to switch current users of new pills to older types of OC or to preferentially prescribe second generation versus third generation OCs.

Acknowledgments

We thank the Edwards–Steptoe Trust for a grant to cover some expenses connected with the preparation of this publication. We also thank Lothar Heinemann for his immensely constructive comments on our paper, and also the many authors of articles in this symposium for their criticisms of the near-final paper.

References


