Epidemiology / Épidémiologie

Observation versus intervention in the evaluation of drugs: the story of hormone replacement therapy

Dominique Costagliola [a,b]

[a] U720, INSERM, BP 335, 75625 Paris cedex 13, France
[b] UMR S 720, Université Pierre-et-Marie-Curie (Paris-6), BP 335, 75625 Paris cedex 13, France

Abstract

Hormone replacement therapy, which was approved for menopausal symptoms, offers an opportunity to compare clinical trials and observational studies when evaluating the risks and benefits of drugs. The differences between randomized and observational evidence relate mainly to the risks of coronary heart diseases and dementia, higher or not elevated in users in trials, and decreased in observational studies. The most likely explanation for these discrepancies is bad accounting for confounders, in particular, time-dependent confounders in classical multivariate analyses and use of prevalent user design. Marginal structural models and new user design should help to diminish strongly indication bias in future observational studies aiming at the evaluation of the risks and benefits of drugs. *To cite this article: D. Costagliola, C. R. Biologies 330 (2007).*
© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

Études observationnelles versus essais dans l'évaluation des traitements : la saga de l'hormonothérapie substitutive de la ménopause. L'hormonothérapie substitutive de la ménopause, parce qu'elle a d'abord été approuvée pour le traitement des troubles climatériques qui surviennent chez certaines femmes lors de la ménopause, offre un cadre pour comparer les essais cliniques et les études observationnelles dans l'évaluation du bénéfice et du risque des traitements médicamenteux. Les différences entre l'évaluation à partir des essais et à partir des études observationnelles portent sur le risque de maladies coronaires et le risque de démence, augmenté ou peu affecté chez les utilisatrices dans les essais, et diminué de façon très importante dans les études d'observation. L'explication la plus plausible de ces différences est une mauvaise prise en compte des facteurs de confusion, notamment ceux dépendant du temps, par les techniques classiques d'analyse multivariée et du choix d'inclure des utilisatrices prévalentes. Les modèles marginaux structuraux et les inclusions de nouveaux utilisateurs uniquement devraient contribuer dans le futur à une meilleure prise en compte des biais d'indication dans les études observationnelles visant à évaluer les risques et les bénéfices d'un traitement. *Pour citer cet article : D. Costagliola, C. R. Biologies 330 (2007).*
© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Observational studies; Clinical trials; Confounding factors; Bias; Hormone replacement therapy; Menopause

Mots-clés : Études observationnelles ; Essais cliniques ; Facteurs de confusion ; Biais ; Thérapie de remplacement hormonal ; Ménopause

E-mail address: dcostagliola@ccde.chups.jussieu.fr.

1. Introduction

Hormone replacement therapy (HRT) was initially approved for the treatment of menopausal symptoms such as hot flashes and night sweats. Then numerous observational studies evaluated the role of HRT on various morbidities such as breast cancer, endometrial cancer, colorectal cancer, strokes, deep vein thrombosis and pulmonary embolism, coronary heart diseases and osteoporosis-related fractures. A striking example of the literature in the domain was the paper published in 1992 in the *Annals of Internal Medicine*, entitled "Hormone therapy to prevent disease and prolong life in postmenopausal women" [1], a title that one may find more suitable for a newspaper article than in a scientific journal. The Women's Health Initiative clinical trial and observational study [2] was a large and complex clinical investigation of strategies of prevention and control of some of the most common causes of morbidity and mortality among postmenopausal women initiated in 1992. The clinical trial was designed to test three distinct interventions, a low-fat eating pattern, hormone replacement therapy, and calcium and vitamin D supplementation. Here we will focus on the HRT intervention only. It was hypothesized that HRT will reduce the risk of coronary heart disease and other cardiovascular diseases and, secondarily, the risk of hip and other fractures, with increased breast cancer risk as a possible adverse outcome. The trial had two parts, one relative to 16 608 postmenopausal women, 50–79 years of age with an intact uterus, comparing conjugated equine oestrogen plus progestin to placebo [3], and one relative to 10 739 postmenopausal women, 50–79 years of age with hysterectomy, comparing the effect of conjugated equine oestrogen to placebo [4]. The first one was interrupted prematurely in May 2002 because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse event and the global index statistic supported risks exceeding benefits after an average 5.2 year of follow-up (minimum 3.5 years, maximum 8.5 years). The second one was also stopped prematurely in March 2004 after an average follow-up of 6.8 years (minimum 5.7 years, maximum 10.7 years) because of the likelihood that neither cardioprotection nor breast cancer risk would be demonstrated in the remaining intervention period, while the excess risk of stroke was similar to the risk reported in the first trial, which was unacceptable for a prevention trial.

The topic of HRT offers then a unique opportunity to assess the role of observational studies in drug evaluation as compared to randomized clinical trials in a field where, for clinical outcomes, observational data preceded clinical trials.

2. Bias and confounding

The usefulness of observational studies in the evaluation of drugs is not the same if the aim of the analysis concerns adverse effects or efficacy. Most people would admit the use of observational studies for evaluating the risk of rare adverse events of drugs, because in that case there is a smaller likelihood of prescription bias linked with good prognosis [5] and reject its use for evaluating the efficacy [6–8].

The reason for this is that well-conducted clinical trials are less prone to bias than observational studies. Three main categories of bias can be defined: selection bias, classification bias, and confounding bias [9]. A selection bias can occur when a participant is lost to follow-up in a cohort study, or when the endpoint of interest cannot be evaluated, or when only prevalent cases are recruited in a case-control study. It may also occur when prevalent users are enrolled instead of new users in a cohort study and the risk varies with time [10]. The healthy worker effect in environmental research or the healthy user effect in pharmacoepidemiology is also an example of selection biases. Classification bias may occur when the person in charge of data collection knows the status of the subject. In cohort studies, classification bias may occur when there is a problem in the detection of the outcome of interest or when there is an error in the status of the subject (with or without the outcome of interest). In case-control studies, a classification bias may occur because case and control do no report exposure with similar reliability, or when there is an error on the measurement of the exposure status. Finally, in an observational setting, a confounding bias can occur when a factor is a risk factor for the outcome of interest, and is associated to exposure too. If the two groups compared differed by the frequency of the confounding factor, this may biased the result of the study. A confounding factor may be accounted for by design, using matching, or by analysis, but it is always difficult to warrant than no confounding remains.

In a well-conducted randomized controlled clinical trial, the treatment is randomized between groups. This procedure controls for selection bias and confounding at enrolment. If the trial is blinded, it also controls for differential measurement errors in the outcome of interest, therefore reducing classification bias. However, some selection bias and confounding can still occur if patients are lost to follow-up or if the outcome of interest cannot be assessed on all subjects enrolled.

However, in the real world, trials may be subject to bias, unreliable (too small, low quality, unpublished), not available, not appropriate, not ethical or not possible. For instance, the cost of the WHI was estimated as $628 million over the 15-year period 1992–2007.

3. What was known prior to the WHI publications?

The best available knowledge at time of the WHI publication was nicely summarized by Grodstein et al. [11]. For hip fracture, a meta-analysis from 1992 [1] estimated the relative risk in women exposed to HRT as 0.75 (95% CI: 0.68–0.84). For coronary heart diseases (CHD), three meta-analyses were available (Table 1); all found a protective effect of HRT on the risk of CHD [1,12,13]. These three meta-analyses were strongly influenced by the results from the Nurse's Health Study (NHS). The Nurses' Health Study began in 1976 when 121 700 female nurses 30 to 55 years of age completed a mailed questionnaire including questions about their postmenopausal hormone use and med-

ical history, including cardiovascular disease and its risk factors. In 1976, women were asked about use and duration of hormone therapy after menopause. Beginning in 1978, information on the types of hormones taken was collected. All information was updated biennially. In 1976, 21 947 postmenopausal women entered the analysis, and 48 586 women were added during follow-up as they became postmenopausal, for a total of 70 533 participants; 808 825 person-years of follow-up were accrued in the most recent updated publication on this topic before the publication of the WHI trial [14]. The NHS investigators estimated the relative risk of CHD at 0.61 (95% CI: 0.52–0.71), with a relative risk of 0.55 (95% CI: 0.45 to 0.68) for women taking oestrogen alone and 0.64 (95% CI: 0.49 to 0.85) for those taking oestrogen plus progestagen. The same paper reported the relative risk for stroke, with a relative risk between stroke and use of oral conjugated oestrogen alone estimated as 1.18 (95% CI: 0.95 to 1.46), and a 45% higher risk for stroke among women taking oestrogen combined with progestin than in those who had never taken hormone therapy (relative risk, 1.45 [95% CI: 1.10 to 1.92]). For pulmonary embolism, the NHS reported a risk of 2.1 (95% CI: 1.2 to 3.8) [15], and similar findings were observed in a large population-based case-control study [16] for pulmonary embolism and deep vein thrombosis, with a risk of 2.1 (95% CI: 1.4 to 3.2). For breast cancer (Table 2), a large meta-analysis [17] reported a relative risk of 0.99 ± 0.08 for oestrogen-only users for less than five years and 1.15 ± 0.17 for oestrogen-plus-progestagen users for less than five years. For use longer or equal to five years, the respective risks were estimated as 1.34 ± 0.09 and 1.53 ± 0.33. For colorectal cancer, a meta-analysis [18] reported a relative risk for HRT users of 0.66 (95% CI: 0.59–0.74).

Table 1
Summary of relative risks for coronary heart disease and hormone replacement therapy in four meta-analyses

Study	RR (95% CI)
Humphrey L 2002[*]	0.87 (0.82–1.21)
	Adjusted on SES[+]
	1.07 (0.79–1.48)
Barret-Connor 1998	0.70 (0.65–0.75)
Grady 1992	0.65 (0.59–0.71)
Stampfer 1991	0.56 (0.50–0.61)

[*] Excluding poor quality, angiography and cross-sectional design studies.
[+] Similar findings when stratifying on studies that adjusted for alcohol consumption or exercise.

Table 2
Summary of relative risks for breast cancer and hormone replacement therapy in a large meta-analysis and two recent studies

Study	Oestrogen only RR (95% CI) or $\pm$SE	Oestrogen plus progestagen RR (95% CI) or $\pm$SE
Fournier et al. 2005	1.1 (0.8–1.6)	Micronized progesterone 0.9 (0.7–1.2)
		Progesterone or testosterone derivatives 1.4 (1.2–1.7)
Million Women Study Collaborators, 2003	Overall: 1.30 (1.22–1.38)	Overall: 2.00 (1.91–2.09)
	<5 years: 1.21 (1.07–1.37)	<5 years: 1.70 (1.56–1.86)
	>5 years: 1.34 (1.23–1.40)	>5 years: 2.21 (2.06–2.36)
Collaborative Group on Hormonal Factors in Breast Cancer, 1997	<5 years: 0.99 ± 0.08	<5 years: 1.15 ± 0.17
	>5 years: 1.34 ± 0.09	>5 years: 1.53 ± 0.33

Finally, for dementia, a meta-analysis [19] reported a risk reduction of 29% (relative risk 0.71, 95% CI: 0.53–0.96). Overall, benefits were found for fracture, colorectal cancer, coronary heart diseases, and dementia – while there was an estimated increased risk of stroke, venous thromboembolic diseases –, and for long exposure to breast cancers.

4. WHI results

The results of both parts of the HRT WHI trial are presented in Table 3. For some endpoints, detailed papers are available for both trial parts (fracture [20,21], coronary heart disease [22,23], stroke [24,25], dementia [26,27], and breast cancer [28,29]); in that case, the results reported here come from these papers. For other endpoints, the oestrogen only trial is not published in a detailed manner (venous thromboembolic disease and colorectal cancer). In that case, I reported the values of the main paper for both trials [3,4].

In terms of benefit, exposure to HRT reduced the risk of fractures, including hip fractures, similarly with oestrogen only and with oestrogen plus progestin. For colorectal cancer, there was a beneficial effect with oestrogen plus progestin, but not with oestrogen only. The risk of ischemic stroke was increased similarly with oe-

Table 3
Main results of the WHI trials

Outcomes	Conjugated equine estrogen plus progestin versus placebo ($N = 16\,608$) Average follow-up: 5.2 yr		Conjugated equine estrogen versus placebo ($N = 10\,739$) Average follow-up: 6.8 yr	
	Hazard ratio	95% CI	Hazard ratio	95% CI
Fractures	0.76	0.69–0.83	0.71	0.64–0.80
Hip	0.67	0.47–0.96	0.65	0.45–0.94
Coronary heart diseases	1.24	1.01–1.54	0.95	0.79–1.16
Year 1 of follow-up	1.81	p for trend 0.02 p for interaction 0.36	1.11	p for trend 0.14 p for interaction 0.07
Age				
50–59 yr	1.27		0.63	
60–69 yr	1.05		0.94	
70–79 yr	1.44		1.11	
Years since menopause or hysterectomy		0.33		0.06
<10	0.89		1.13	
10–19	1.22		0.66	
⩾20	1.71		1.09	
Stroke	1.31	1.02–1.68	1.37	1.09–1.73
Ischaemic	1.44	1.09–1.90	1.55	1.19–2.01
Haemorrhagic	0.82	0.43–1.56	0.64	0.35–1.18
Invasive breast cancer	1.24	1.01–1.54	0.80	0.62–1.04
Dementia[*] ($n = 4532$ and 2947, respectively)	2.05	1.21–3.48	1.49	0.83–2.66
Colorectal cancer	0.63	0.43–0.92	1.08	0.75–1.55
Venous thromboembolic disease	2.11	1.58–2.82	1.33	0.99–1.79
Deep vein thrombosis	2.07	1.49–2.87	1.47	1.04–2.08
Pulmonary embolism	2.13	1.39–3.25	1.34	0.87–2.06
Death	0.98	0.82–1.18	1.04	0.88–1.22

[*] Pooled estimate 1.76 (1.19–2.60).

strogen only and with oestrogen plus progestin. The risk of venous thromboembolic disease was increased similarly for deep vein thrombosis and pulmonary embolism, with a slightly higher risk with oestrogen plus progestin than with oestrogen only. The risk of breast cancer was significantly increased only with oestrogen plus progestin, but not with oestrogen. For coronary heart disease, there was an increased risk with oestrogen plus progestin, but not with oestrogen. The risk was increased in the first year of use with oestrogen plus progestin but not with oestrogen. No trends were evidenced for age at enrolment or years since menopause in the oestrogen plus progestin trial. For the oestrogen alone trial, there was a trend with the age at enrolment, with a risk estimated at 0.63 in women aged 50–59 years, 0.94 in women aged 60–69 years and 1.11 in women aged 70–79 years. For years since hysterectomy, there was also a trend, which was not linear (1.13 for women enrolled less than 10 years after hysterectomy, 0.66 for women enrolled less than 10–19 years after, and 1.09 for women enrolled more than 20 years after). Finally, for dementia, in a sub-trial limited to women above 65 years of age without dementia at baseline, the pooled estimate of risk in both parts of the trial was 1.76 (95% CI: 1.19–2.60).

Compared to what was known, results from the oestrogen-plus-progestin trials were surprising for CHD and dementia. The results of the oestrogen trial were less commented. No effect was seen for CHD in this trial. Surprisingly, no increased risk of breast cancer was evidenced, and there was even a trend toward a reduced risk.

5. Can we resolve the discrepant results between observational evidence and WHI for CHD?

A meta-analysis submitted prior to the first publication of the oestrogen plus progestin WHI trial, but published after, re-examined the evidence available on the risk of CHD [30]. In this paper, studies of poor quality, studies with angiography endpoints, and cross-sectional design studies were excluded. Further, the authors found that a cause of heterogeneity between the studies was linked with whether or not the analysis was adjusted on socio-economic status, alcohol consumption, or exercise. Their pooled estimate (Table 1) was 0.87 (0.82–1.21) overall and 1.07 (0.79–1.48) when considering only studies that were adjusted on socio-economic status. For instance, in a good-quality case-control study [31], in which the analysis was adjusted on socio-economic status and alcohol consumption, the estimated relative risk was 0.9 for ever use of unopposed

oestrogen (95% CI: 0.7–1.2), and there were insufficient data to evaluate oestrogens taken together with progestagen.

Other confounding effects may also have played a role. Up to the mid 1990s, HRT was contraindicated for women with hypertension, diabetes, and cardiovascular diseases and therefore opening the door to a huge indication bias, unlikely to be well controlled by usual adjustment techniques in most observational studies. In addition, HRT users tend to be healthier than non-users, more educated, with higher social class, leaner, with more positive health behaviours, better levels of several cardiovascular risk factors and better compliance. Better compliance to drugs is a well-known factor of good prognosis.

In WHI, the increased risk of CHD in the oestrogen-plus-progestin treatment was concentrated in the initial year after starting hormones. In the NHS, exposure data were updated biennially. Therefore, duration of exposure was underestimated by one year and this may have contributed to the miss of the early increased risk. In addition, women having already gone through the menopause in 1976 were considered in the analysis, which therefore included prevalent users in addition to incident users. As pointed out by Ray [10], this practice can cause two types of bias, both of which plausibly may have contributed to the discrepancy between observational and randomized studies. First, prevalent users are 'survivors' of the early period of pharmacotherapy, which can introduce substantial bias if risk varies with time, just as in studies of operative procedures that enrol patients after they have survived surgery. Second, covariates for drug users at study entry often are plausibly affected by the drug itself, which may introduce confounding. To account for the impact of time-dependent confounding factors, which is linked with the probability of initiating HRT and with the risk of CHD or of other endpoints and is also affected by HRT, and therefore intermediate on the causal pathway from HRT to the endpoint, the usual approaches, such as stratification or regression models may lead to biased results. To overcome this problem, Robins et al. introduced marginal structural models [32,33]. First, the probability of each subject to be treated at each time is estimated, using logistic regression. Second, the weight of each subject, defined as the inverse of each subject's probability of his or her treatment history at each time is derived. Then, a weighted Cox model is fitted for the effect of treatment, controlling for baseline covariates, but not for time-dependent confounders. The weighted analysis creates a statistical population in which the probability of being treated at each time is

unrelated to the measured prognostic factors (the time-dependent confounders). The validity of the method depends on all prognostic factors of treatment initiation being recorded. J. Robbins presented a reanalysis of the NHS using this approach in 2004 and showed that the results were then much closer to that of the WHI trial for the risk of CHD [34].

Some authors argued that women who received HRT in real life are much younger that women enrolled in the WHI HRT trials. However, the analysis by age at enrolment and by time since menopause or hysterectomy (Table 2) did not support the view that this might explain the difference. Finally, in sub-group analysis reported in [22], no interaction was seen with the factors studied, including CHD risk factors. In others words, the increased risk was the same in women with and without risk factors, including overweight. These results dismiss the argument of those who consider that we cannot use the results of the WHI trial for women with lower numbers of CHD risk factors.

Of note, a common belief that was used to justify the protective effect of HRT is that risk in women is held low only until the menopause, around age 50 years, when it rebounds, equalling, and later surpassing that in men [35]. This is just not supported by mortality/morbidity data, as shown by Tunstall-Pedoe [36], who evidenced that the risk of death from CHD by age in the United Kingdom is parallel in men and women. Myocardial infarction incidence data from the three French Monica registries in France exhibited a similar trend. There is no change in the age incidence trend in women around or after 50 years of age, and relative risks in women compared to men are similar around 35–39 years (4.7), an age when almost no women has gone though the menopause, and around 60–64 years (4.4), an age when almost all women have, as can be seen in Fig. 1.

Overall, most of the difference between observational studies and WHI trials can be explained by default in analysis such as omission of confounders, bad adjustment for time-dependent confounders, and no good account of the function of risk overtime since initiation of treatment in the design and analysis of the studies. The performed analyses were influenced by the belief that the impact of HRT on CHD was mainly mediated by their positive impact on the lipid profiles, therefore the early impact of HRT was not strongly considered and accounted for in many analyses. It is also quite likely that similar problems explained the difference seen between observational studies and the two WHI HRT trials for dementia risk.
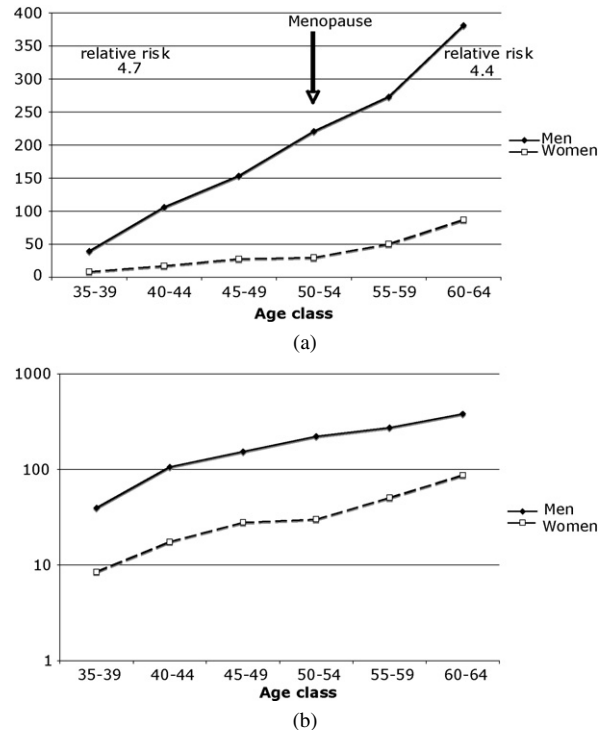

Fig. 1. Incidence of myocardial infarction by sex and age (**a**) and on a log scale (**b**) in France in 1999–2001 (Monica-France).

This does not mean that we know all what is needed to take decision as to whether or not use HRT. For instance, the transdermal route of oestrogen administration could have a different impact on the CHD or ischemic stroke risks, as is plausible for venous thromboembolism [37], and deserves further studies.

6. What commentary can be made on breast cancer risk?

Since the publication of the first WHI trial, a large study, the Million Women Study (MWS), reported the risk of HRT for breast cancer [38]. This study enrolled women attending mammography screening who fulfilled a questionnaire relative to HRT exposure and potential confounders. The cancer status was evaluated through cancer registries covering the UK population. Because of its size (over 800 000 women included in the analysis), this study allowed one to study the role of different oestrogens and progestagens, as well as different routes of administration. This study, as the WHI trials, showed that the risk associated with exposure to oestrogen plus progestagen is higher than the risk of oestrogen-only regimen. However, the risks were slightly higher in MWS than in the WHI trials (Table 2). For instance, the risk was 1.24 overall in the E+P WHI

trial, while it was 2.00 in the MWS. One may however wonder if the best summary of a trial is the relative risk from the intent to treat (ITT) analysis when considering a harmful effect. One may hypothesize that the observational estimates are closer to the on-treatment risk than to the ITT risk. The on-treatment risk for the E+P WHI trial is 1.49 (instead of 1.24 for the ITT risk). On the other hand, for the E WHI trial, no risk was evidenced, and even there was a trend to a beneficial effect. In that case, it might be more prudent and more conservative to keep the ITT risk (0.80) rather than the on-treatment risk (0.67). These results are nonetheless at variance with the estimated increased risk in the MWS (RR of 1.3 for oestrogen-only exposure). Therefore, it remains unclear whether or not the risk is truly augmented after exposure to oestrogen only.

In the MWS study, in oestrogen-only users, the risks were similar by constituent, by dose and by route of administration. Similarly, in oestrogen-plus-progestagen users, no difference was found according to the progestagen constituent or the regimen type (sequential or continuous). One progestagen, which is in common use in France (micronized progesterone), was evaluated neither in the WHI trials nor in the MWS or in recent literature in Nordic countries. In fact, there is only one study reporting separates estimates for micronized progesterone versus progesterone or testosterone derivatives [39]. One may first note that this study, E3N, is the only one using the new users' design [10], and is therefore more likely to be less prone to the confounding described in Section 2 of this paper, although limited at the moment by a short mean follow-up given the new user design (5.8 years). In this study, as shown in Table 2, the risk for exposure to oestrogen only was 1.1 (95% CI: 0.8–1.6) for a mean exposure duration of 2.4 years. For exposure to oestrogen plus progestagen, there was a difference according to the progestagen type, the risk for oestrogen plus micronized progesterone was estimated as 0.9 (95% CI: 0.7–1.2) after a mean exposure of 3.0 years, while for oestrogen plus either progesterone or testosterone derivatives, it was estimated as 1.4 (95% CI: 1.2–1.7) after a mean exposure of 2.9 years. It should be noted that these results are prettily similar to those of the WHI trial, apart for the risk of micronized progesterone, which was not evaluated in the WHI trials, which is quite reassuring for the quality of analysis and design to limit bias and confounding in this study. Overall, with only one relatively short-term study evaluating micronized progesterone, it is difficult to conclude firmly on the potential difference according to the progestagen type with available data, and on whether a less risky regimen can be rec-

ommended. Micronized progesterone is also in use in Belgium, Germany, Italy, Spain, and Switzerland, and, hopefully, studies from these countries will help understanding the E3N results.

Of note, both in the WHI trials and in the MWS, breast cancer risks were independent of age and body mass index. Therefore, for this endpoint also, arguments tuning down the results of these studies based on differences between real-life HRT users in France are not scientifically based.

Finally, recent ecological evidence in the USA showed both a dramatic fall in HRT use and a subsequent decline in breast cancer incidence [40], and data from other countries are awaited to confirm this decline.

7. Conclusion

HRT does not prevent many diseases and prolong life, but is efficacious for menopausal symptoms. A HRT, which could potentially be associated with a much lower overall risk such as transdermal route for oestrogen and micronized progesterone, still needs to be properly evaluated. In addition, the fact that, in the E+P WHI trial, all risks were at least similar or higher than those observed in the E WHI trial, apart from the risk of colorectal cancer, raised some concerns on the safety of progestagen use, namely progestagen used during perimenopause, a well-developed practice in France.

On a more general level, observational data should not replace clinical trials. Cohort studies and clinical trials have complementary roles in the evaluation of treatment. Cohort studies are very useful to assess rare long-term adverse effects of treatment that cannot be evaluated in clinical trials. The main problem with using cohort data to evaluate the effect of treatments is linked with confounding by indication and it is therefore critical to evaluate thoroughly the care taken to account for bias in observational data design and analyses. While validating the proper analysis of observational data by comparison with trial results is of interest, the main focus of such analyses should be on questions not answered, or not answerable by trials. Some recently presented methods, such as marginal structural models and new user design, may help to produce less biased estimates of the treatment effect from observational data, particularly in cases of time-dependent confounding, when no trial data is available.

Acknowledgements

References

[1] D. Grady, S.M. Rubin, D.B. Petitti, C.S. Fox, D. Black, B. Ettinger, V.L. Ernster, S.R. Cummings, Hormone therapy to prevent disease and prolong life in postmenopausal women, Ann. Intern. Med. 117 (1992) 1016–1037.

[2] The Women's Health Initiative Study Group, Design of the women's health initiative clinical trial and observational study, Contr. Clin. Trials 19 (1998) 61–109.

[3] J.E. Rossouw, G.L. Anderson, R.L. Prentice, A.Z. LaCroix, C. Kooperberg, M.L. Stefanick, R.D. Jackson, S.A. Beresford, B.V. Howard, K.C. Johnson, J.M. Kotchen, J. Ockene, Writing Group for the Women's Health Initiative Investigators, Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial, J. Am. Med. Assoc. 288 (2002) 321–333.

[4] G.L. Anderson, M. Limacher, A.R. Assaf, T. Bassford, S.A. Beresford, H. Black, D. Bonds, R. Brunner, R. Brzyski, B. Caan, R. Chlebowski, D. Curb, M. Gass, J. Hays, G. Heiss, S. Hendrix, B.V. Howard, J. Hsia, A. Hubbell, R. Jackson, K.C. Johnson, H. Judd, J.M. Kotchen, L. Kuller, A.Z. LaCroix, D. Lane, R.D. Langer, N. Lasser, C.E. Lewis, J. Manson, K. Margolis, J. Ockene, M.J. O'Sullivan, L. Phillips, R.L. Prentice, C. Ritenbaugh, J. Robbins, J.E. Rossouw, G. Sarto, M.L. Stefanick, L. Van Horn, J. Wactawski-Wende, R. Wallace, S. Wassertheil-Smoller, Women's Health Initiative Steering Committee, Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial, J. Am. Med. Assoc. 291 (2004) 1701–1712.

[5] J.P. Vandenbroucke, When are observational studies as credible as randomised trials?, Lancet 363 (2004) 1728–1731.

[6] J. Concato, N. Shah, N.J. Horwitz, Randomized, controlled trials, observational studies, and the hierarchy of research designs, N. Engl. J. Med. 342 (2000) 1887–1892.

[7] D.A. Lawlor, G. Davey Smith, K.R. Bruckdorfer, D. Kundu, S. Ebrahim, Observational versus randomised trial evidence, Lancet 363 (2004) 1724–1727.

[8] S. MacMahon, R. Collins, Reliable assessment of the effects of treatment on mortality and major morbidity. II: Observational studies, Lancet 357 (2001) 455–462.

[9] J. Bouyer, D. Hémon, S. Cordier, F. Derriennic, I. Stüker, B. Stengel, J. Clavel, Les biais, in: J. Bouyer, D. Hémon, S. Cordier, F. Derriennic, I. Stüker, B. Stengel, J. Clavel (Eds.), Epidémiologie, principes et méthodes quantitatives, Éditions INSERM, Paris, 1995, pp. 89–156.

[10] W.A. Ray, Evaluating medication effects outside of clinical trials: new-user designs, Am. J. Epidemiol. 158 (2003) 915–920.

[11] F. Grodstein, T.B. Clarkson, J.A.E. Manson, Understanding divergent data on postmenopausal hormone therapy, N. Engl. J. Med. 348 (2003) 645–650.

[12] M.J. Stampfer, G.A. Codditz, Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence, Prev. Med. 20 (1991) 47–63.

[13] E. Barrett-Connor, D. Grady, Hormone replacement therapy, heart disease, and other considerations, Annu. Rev. Public Health 19 (1998) 55–72.

[14] F. Grodstein, J.E. Manson, G.A. Colditz, W.C. Willett, F.E. Speizer, M.J. Stampfer, A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease, Ann. Intern. Med. 133 (2000) 933–941.

[15] F. Grodstein, M.J. Stampfer, S.Z. Goldhaber, J.A.E. Manson, G.A. Colditz, F.E. Speizer, W.C. Willett, C.H. Hennekens, Prospective study of exogenous hormones and risk of pulmonary embolism in women, Lancet 348 (1996) 983–987.

[16] S. Pérez-Guttann, L.A. Garcia Rodríguez, J. Castellsague, A. Duque Oliart, Hormone replacement therapy and the risk of venous thromboembolism: population based case-control study, Brit. Med. J. 314 (1997) 796–800.

[17] Collaborative Group on Hormonal Factors in Breast Cancer, Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer, Lancet 350 (1997) 1047–1059.

[18] F. Grodstein, P.A. Newcomb, M.J. Stampfer, Postmenopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis, Am. J. Med. 106 (1999) 574–582.

[19] K. Yaffe, G. Sawaya, I. Lieberburg, D. Grady, Estrogen therapy in postmenopausal women: effects on cognitive function and dementia, J. Am. Med. Assoc. 279 (1998) 688–695.

[20] J.A. Cauley, J. Robbins, Z. Chen, S.R. Cummings, R.D. Jackson, A.Z. LaCroix, M. LeBoff, C.E. Lewis, J. McGowan, J. Neuner, M. Pettinger, M.L. Stefanick, J. Wactawski-Wende, N.B. Watts, Women's Health Initiative Investigators, Effects of estrogen plus progestin on risk of fracture and bone mineral density: the Women's Health Initiative randomized trial, J. Am. Med. Assoc. 290 (2003) 1729–1738.

[21] R.D. Jackson, J. Wactawski-Wende, A.Z. LaCroix, M. Pettinger, R.A. Yood, N.B. Watts, J.A. Robbins, C.E. Lewis, S.A. Beresford, M.G. Ko, M.J. Naughton, S. Satterfield, T. Bassford, Women's Health Initiative Investigators, Effects of conjugated equine estrogen on risk of fractures and BMD in postmenopausal women with hysterectomy: results from the women's health initiative randomized trial, J. Bone Miner. Res. 21 (2006) 817–828.

[22] J.E. Manson, J. Hsia, K.C. Johnson, J.E. Rossouw, A.R. Assaf, N.L. Lasser, M. Trevisan, H.R. Black, S.R. Heckbert, R. Detrano, O.L. Strickland, N.D. Wong, J.R. Crouse, E. Stein, M. Cushman, Women's Health Initiative Investigators, Estrogen plus progestin and the risk of coronary heart disease, N. Engl. J. Med. 349 (2003) 523–534.

[23] J. Hsia, R.D. Langer, J.E. Manson, L. Kuller, K.C. Johnson, S.L. Hendrix, M. Pettinger, S.R. Heckbert, N. Greep, S. Crawford, C.B. Eaton, J.B. Kostis, P. Caralis, R. Prentice, Women's Health Initiative Investigators, Conjugated equine estrogens and coronary heart disease: the Women's Health Initiative, Arch. Intern. Med. 166 (2006) 357–365.

[24] S. Wassertheil-Smoller, S.L. Hendrix, M. Limacher, G. Heiss, C. Kooperberg, A. Baird, T. Kotchen, J.D. Curb, H. Black, J.E. Rossouw, A. Aragaki, M. Safford, E. Stein, S. Laowattana, W.J. Mysiw, WHI Investigators, Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative: a randomized trial, J. Am. Med. Assoc. 289 (2003) 2673–2684.

[25] S.L. Hendrix, S. Wassertheil-Smoller, K.C. Johnson, B.V. Howard, C. Kooperberg, J.E. Rossouw, M. Trevisan, A. Aragaki, A.E. Baird, P.F. Bray, J.E. Buring, M.H. Criqui, D. Herrington, J.K. Lynch, S.R. Rapp, J. Torner, WHI Investigators, Effects of conjugated equine estrogen on stroke in the Women's Health Initiative, Circulation 113 (2006) 2425–2434.

[26] S.A. Shumaker, C. Legault, S.R. Rapp, L. Thal, R.B. Wallace, J.K. Ockene, S.L. Hendrix, A.R. Jones, B.N. Assaf 3rd, R.D. Jackson, J.M. Kotchen, S. Wassertheil-Smoller, J. Wactawski-

Wende, WHIMS Investigators, Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in post-menopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial, J. Am. Med. Assoc. 289 (2003) 2651–2662.

[27] S.A. Shumaker, C. Legault, L. Kuller, S.R. Rapp, L. Thal, D.S. Lane, H. Fillit, M.L. Stefanick, S.L. Hendrix, C.E. Lewis, K. Masaki, L.H. Coker, Women's Health Initiative Memory Study, Conjugated equine estrogens and incidence of proba-ble dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study, J. Am. Med. Assoc. 291 (2004) 2947–2958.

[28] R.T. Chlebowski, S.L. Hendrix, R.D. Langer, M.L. Stefanick, M. Gass, D. Lane, R.J. Rodabough, M.A. Gilligan, M.G. Cyr, C.A. Thomson, J. Khandekar, H. Petrovitch, A. McTiernan, WHI Investigators, Influence of estrogen plus progestin on breast can-cer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial, J. Am. Med. As-soc. 289 (2003) 3243–3253.

[29] M.L. Stefanick, G.L. Anderson, K.L. Margolis, S.L. Hendrix, R.J. Rodabough, E.D. Paskett, D.S. Lane, F.A. Hubbell, A.R. Assaf, G.E. Sarto, R.S. Schenken, S. Yasmeen, L. Lessin, R.T. Chlebowski, WHI Investigators, Effects of conjugated equine estrogens on breast cancer and mammography screening in post-menopausal women with hysterectomy, J. Am. Med. Assoc. 295 (2006) 1647–1657.

[30] L.L. Humphrey, B.K.S. Chan, H.C. Sox, Postmenopausal hor-mone replacement therapy and the primary prevention of Car-diovascular Disease, Ann. Intern. Med. 137 (2002) 273–284.

[31] L. Rosenberg, J.R. Palmer, S. Shapiro, A case-control study of myocardial infarction in relation to use of estrogen supplements, Am. J. Epidemiol. 137 (1993) 54–63.

[32] J.M. Robins, Marginal structural models versus structural nested models as tools for causal inference, in: E. Halloran, D. Berry (Eds.), Statistical Models in Epidemiology: The Environment and Clinical Trials, Springer-Verlag, New York, 1999, pp. 95–134.

[33] J.M. Robins, M.A. Hernán, B. Brumback, Marginal structural models and causal inference in epidemiology, Epidemiology 11 (2000) 550–560.

[34] J.M. Robins, Methodological approaches for reanalyzing ob-servational cohort studies of hormone therapy as a nested se-quence of randomized trials, in: 20th International Conference on Pharmacoepidemiology & Therapeutic Risk Management, 22–25 August 2004, Bordeaux, France.

[35] K.T. Khaw, Epidemiology of the menopause, Br. Med. Bull. 48 (1992) 249–261.

[36] H. Tunstall-Pedoe, Myth and paradox of coronary risk and the menopause, Lancet 351 (1998) 1425–1427.

[37] P.Y. Scarabin, E. Oger, G. Plu-Bureau, EStrogen and THrom-boEmbolism Risk Study Group, Differential association of oral and transdermal oestrogen-replacement therapy with venous thromboembolism risk, Lancet 362 (2003) 428–432.

[38] Million Women Study Collaborators, Breast cancer and hormone-replacement therapy in the Million Women Study, Lancet 362 (2003) 419–427.

[39] A. Fournier, F. Berrino, E. Riboli, V. Avenel, F. Clavel-Chapelon, Breast cancer risk in relation to different types of hormone re-placement therapy in the E3N-EPIC cohort, Int. J. Cancer. 114 (2005) 448–454.

[40] C.A. Clarke, S.L. Glaser, C.S. Uratsu, J.V. Selby, L.H. Kushi, L.J. Herrinton, Recent declines in hormone therapy utilization and breast cancer incidence: clinical and population-based evi-dence, J. Clin. Oncol. 24 (2006) e49–e50.