Court File No.: 05-CV-303001PD2

ONTARIO
SUPERIOR COURT OF JUSTICE

BETWEEN:

CANWEST MEDIAWORKS INC.

Applicant

and

ATTORNEY GENERAL OF CANADA

Respondent

AFFIDAVIT OF JAMES HANLEY

I, **JAMES HANLEY**, of the City of Montreal, in the Province of Quebec, solemnly AFFIRM:

1. I am a Professor in the Departments of Epidemiology & Biostatistics and Medicine at McGill University, a position that I have held since 1980 (as an Associate Professor from 1980 to 1993, and as a Full Professor since then). I served as an Assistant Professor at the Harvard School of Public Health from 1977 to 1980 and the State University of New York from 1973 to 1977. I obtained a PhD in Statistics from the University of Waterloo in 1973, and MSc and BSc degrees in Mathematics and Statistics from the National University of Ireland (University College Cork) in 1969 and 1968.

2. I am currently an Associate Editor for the journal Biometrics, published by the International Biometric Society. Previously, I have been on the

Editorial Board of the journals Statistics in Medicine, Medical Decision Making and Investigative Radiology. I have recently been a Statistical Consultant for the journal Hypertension and Associate Editor for Biostatistics for the Canadian Medical Association Journal. I was for three years a member of the epidemiology grant review panel of the (US) National Institutes of Health (NIH), and an ad hoc member for the NIH diagnostic radiology grant review panel. I have also served on review panels for Canadian and provincial research funding agencies. I spent the 1985-1965 year as a consultant at the Cancer Unit of the WHO in Geneva, and 1992-1993 as a teacher, under the McGill-Ethiopia project, in the Masters of Public Health program at the University of Addis Ababa in Ethiopia.

3. I have authored or co-authored over 200 research articles, both methodological and substantive. Of these, the most relevant to the present task are those dealing with evaluations, based on both observational and experimental study designs, of the efficacy of various forms of treatment for prostate and other cancers, and – using the administrative databases in Saskatchewan -- the benefits and risks of various medications. My most cited articles are those on the methods I developed to statistically evaluate the performance of medical diagnostic tests. I have also authored several expository articles on statistical methods, aimed at statisticians, epidemiologists, public health researchers, and radiologists.

4. My full curriculum vitae is attached to this affidavit as **Exhibit "A"**.

5. I have been asked by counsel for the Attorney General of Canada to provide my opinion on the following:

(a) Issues raised in the Affidavit of Stephen Walter, sworn May 29, 2007, that has been filed by CanWest Global Mediaworks Inc. in this proceeding, about a study by Barbara Mintzes and colleagues entitled "How does direct-to-consumer advertising (DTCA) affect prescribing? A survey in primary care environments with and without DTCA" ("Mintzes Study"). This study was published in the Canadian Medical Association Journal (CMAJ) on September 2, 2003. A shorter version appeared under the title "Influence of direct to consumer pharmaceutical advertising and patients' requests on prescribing decisions: two site cross sectional study" in the British Medical Journal on February 2, 2002;

(b) The methods and results of the Mintzes study.

6. In providing my opinion, I have reviewed the Mintzes Study; the above-cited BMJ article; the reply, in the March 2, 2004 issue of the CMAJ, by Barbara Mintzes and colleagues to the two letters by John Graham, and Marc Lacroix to the Editor of the CMAJ; and the affidavit of Stephen Walter.

7. My evidence and opinions in this affidavit are based on my own knowledge and experience gained through my years of education and experience in the fields of Epidemiology and Biostatistics. My opinion is informed by a critical examination of the items listed above. Where I have quoted the opinions of other authors, they are widely accepted as experts in the matters on which they have written, and their statements reflect my own opinion as well. Finally, I have no conflict of interest in this matter.

A. INTRODUCTION

8. In this affidavit, I will address :

(a) Issues raised about the study and contained in the Affidavit of Stephen Walter, sworn May 29, 2007, that has been filed by CanWest Global Mediaworks Inc. in this proceeding.

(b) My own interpretation of the quality of the methodology of the Mintzes study, the findings in the study, and the interpretations made by its authors.

B. SUMMARY

9. In my opinion, the general comments in the affidavit of Stephen Walter concerning the ideal study, and the extra precautions we must take with observational studies, are correct. However, I emphasize that most of what we know from epidemiologic and public health research, particularly about the undesirable/harmful health effects of certain human behaviours and external agents, we have learned from observational studies on humans.

10. In addition, in my opinion, the objections concerning (1) selection bias; (2) sample size; (3) inaccuracies in patient responses; and (4) statistical adjustment raised in the affidavit of Stephen Walter should be given less weight than he attributes to them. These objections led him to conclude that the Mintzes study "provides at best only weak evidence concerning the effects of DTCA", a conclusion which I believe to be unfounded.

11. As I explain in detail below, (1) no concrete or plausible selection scenarios that would lead to distorted odds ratios were suggested in his affidavit;

(2) the sample size did not need to be as large as he argued it should, since several factors were already quite similar in the compared cities, and the "10:1" criterion he invoked has been shown to be conservative in this type of data-analysis; (3) errors in patient reporting would most likely have attenuated the differences, not inflated them; the fact that statistically significant differences were observed despite these alleged inadequacies in sample size and data quality suggests that the signal would have been even stronger had the sample sizes been larger and the errors smaller; and (4) the authors of the study did carry out statistical adjustments.

12. Thus, since the objections raised cannot be substantiated, I conclude that the Mintzes study does provide evidence about the effects of DTCA. Moreover, the results observed in the Mintzes study have additional plausibility, since they are in line with what one would expect from successful DTCA.

13. In my opinion, the Mintzes study provides evidence that a greater proportion of the patients in the Sacramento than the Vancouver practices studied reported having been exposed to the DTCA drugs studied; it also shows that a greater proportion of Sacramento patients requested these DTCA drugs. Furthermore, within each city separately, a greater proportion of those who reported having seen more DTCA drugs requested these advertised drugs. The authors of the Mintzes study have made a statistically appropriate effort to correct for factors that might have distorted these comparisons. The statistical

analysis took account of the actual degree of similarity in the drug-requesting behaviours of patients seen by the same physician.

14. Quite apart from any causal conclusions one can draw solely from the numerical data in the Mintzes study, one can also infer the consequences of DTCA from first principles, i.e., from the nature and intent of DTCA. Presumably, when DTC advertisements suggest that patients ask their doctor about the DTCA drug, then if these advertisements succeed, a fraction of patients will do so. Likewise, some (unknown) fraction of patient request for these drugs will be honoured.

15. Despite the "coarseness" with which Mintzes and colleagues could measure physicians' opinions concerning prescriptions they wrote, some patterns in the data in their Table 6 are notable. First, both in Sacramento and in Vancouver, physicians were ambivalent (i.e., judged the medicine to be an "unlikely" or "possible" choice, rather than a "likely" choice, for similar patients) about a significantly greater proportion of the prescriptions they wrote in response to patient requests than the ones they wrote when the prescribed drug was not requested by the patient. Second, in Sacramento, the proportion of (patient-requested) instances about which physicians were ambivalent was equally high for DTCA and non-DTCA drugs; Vancouver physicians were ambivalent in a greater proportion of the DTCA than the non-DTCA ones, but the smaller Vancouver denominators do not allow us to precisely quantify the difference. Since physicians appear to be ambivalent in a proportion of the

instances in which they prescribe a drug in response to a patient request, I conclude that DTCA will increase both the prescribing of DTCA drugs in instances where physicians will not be ambivalent and the prescribing of DTCA drugs in instances where they will be ambivalent.

C. ANALYSIS

16. In the following sections I consider the opinions of Professor Walter, and give an expanded version of my own opinions when they differ substantially from his. To make it easier to follow, and since I would have independently used many of the same headings, I use the same numbering system and headings as he did.

1) Observational versus Experimental Studies

a) We can learn from carefully conducted observational studies

17. Although Professor Walter begins, in his paragraph 11, by describing the benefits of the ideal (experimental) study, which randomly assigns some subjects to receive the exposure of interest, and a comparable group to not receive it, he concedes that the nature of DTCA precludes this design. I agree, but I emphasize that one can learn from properly conducted studies that use observational designs.

18. Despite the additional challenges involved in observational studies, much of what we have learned in epidemiologic and public health research,

particularly about the undesirable/harmful health effects of certain human behaviours and external agents, we have learned from observational studies on humans. Examples include the health effects of direct and indirect cigarette smoking, the role of sleeping position in the aetiology of sudden infant death, and the mode of transmission of infectious diseases. In many instances, for ethical or logistic reasons, the etiologic role of these agents and behaviours and the postulated routes of infection could not have been experimentally tested in humans. In some, particularly those in which human cognition is involved, there may be no suitable animal model.

19. The potential benefits of motor cycle helmets, seat belts, speed limits, and health messages and other regulatory actions could be inferred from physical and psychological principles, and tested in laboratory crash tests, and experiments in psychology laboratories. But ultimately the actual benefit is evaluated in the community once the regulations are put in place. Sometimes, we can exploit a natural experiment, such as when a law is introduced, preferably with a staggered timetable, in some jurisdictions and not in others (as when, in response to government prompting, one of the water companies supplying the population of London moved its water intake further upstream; the renowned 19th century British epidemiologist John Snow used this window to demonstrate how cholera was spread).

20. Sometimes, we are necessarily limited to contemporaneous comparisons of persons in different jurisdictions with different

policies/regulations. Before cable and satellite television, researchers were able to study the effects of the content of television programs and health messages by comparing the behaviour of sub populations living in the same state, but isolated by mountains and other barriers from different television channels. The world has become smaller and thus the study units are necessarily larger, but we will continue to learn from carefully performed observational studies on these units.

b) Mintzes et al. did deal with the possibility of confounding

21. At the end of paragraph 12 of his affidavit, Professor Walter correctly points out that the observational nature of the Mintzes study requires that the authors carefully consider the possibility of confounding of the effect of DTCA exposure on patient and physician outcomes. I contend that Mintzes et al. did in fact do a reasonable job in dealing with this possibility, and give my reasons below.

22. Epidemiologists use the term "confounding" for the situation where the comparison of outcomes in those exposed/not exposed to the factor of concern is distorted by the fact that those exposed and not exposed differ with respect to another (nuisance) factor that itself influences the outcome. Chapter 1 of the 2002 text by K. Rothman[1] gives a vivid and real didactic example: A 20-year follow-up of UK women, first surveyed in the 1970s, found that a far greater proportion of the non-smokers than the smokers had died (smoking is the factor of interest). The confounding factor was the age of the women: the non-smokers

on average were older, from an era when smoking among women was less common, whereas the smokers tended to be from the younger generation. This difference in age, a key determinant of mortality rates, distorted the comparison, and led to the artefact epidemiologists call confounding (confused effects). This artefact disappears (and the ratio of the death rates is reversed) once one adjusts, in the data analysis, for the difference in the age profiles of the smokers and non-smokers, i.e., once one "compares like with like."

c) *In observational studies, not every determinant/difference is a confounder.*

23. In paragraph 13 of his affidavit, Professor Walter refers to several "determinants" of patient-requesting and physician-prescribing of drugs, and lists several potential ones. The list should not be taken to mean that every one of these needs to be included in the matching or regression models that Mintzes et al. employed. In observational studies, not every determinant/difference is a confounder. One does not have to necessarily adjust for those variables that (a) influence the target response, but have similar profiles in the compared groups; (b) do not influence the target response, but have different profiles in the compared groups.

24. Thus, there is a difference between (i) a potential determinant (the list could be very long), (ii) a potential confounder (likewise), (iii) an actual but unmeasured confounder and (iv) an actual but measured confounder one can

---

[1] Rothman KJ Epidemiology: An Introduction, Oxford University Press, 2002.

adjust for. It is the number and strength of the variables in (iii) that matters. The preferred way to minimize bias from unmeasured and unrecognized confounding variables is to employ random assignment of subjects to the compared exposures, but this option is not available in this situation. Therefore, just as is done in all of observational epidemiology, the only recourse is to think hard about what variables could confound, to measure them, and adjust for them as necessary. Mintzes et al considered several possible confounding variables: Table 1 lists a large number of candidates; they adjust for many of them in the analyses shown in Tables 3 to 6.

25. Incidentally, statistical adjustment is sometimes necessary (i.e., there can be confounding) even in randomized experiments. A striking example is the elegant experimental study[2], on how exposure to scientific theories about women's math abilities affects women's math performance. Despite the random allocation of the 133 women to the compared exposures, statistical adjustment was required in order to adjust for an important difference in the pre-exposure profiles of the compared exposure groups.

d) Some of the reported differences between the compared groups in the Mintzes study are not that strong/important; and one factor, for which there is evidence that it tends to reduce prescribing, was more common in the Sacramento sample.

26. In his paragraph 15, Professor Walter refers to the many other differences that exist between those exposed and not exposed to DTCA, and

contends that statistical analyses require large samples to deal with such "substantial" differences and to isolate the particular effect of DTCA.

27. As discussed below, some of the reported differences between the compared groups are not that strong, or even go in the opposite effect from what was observed. Even those variables that do show substantial differences, and were corrected for in the statistical analysis, collectively do not appear to have greatly affected the results of comparison of interest. Moreover, in their interpretation (lines 7-9 of page 412 of the CMAJ article), Mintzes et al. argue, correctly in my opinion, that, if anything, the fact that most Sacramento physicians in the study were salaried, would mean they would be *less* likely to be influenced by incentives to prescribe. They base their reasoning on a study that shows that such incentives tend to be greater under a fee for service system.

2) Selection Biases

a) Selectivity in and of itself does not necessarily invalidate comparisons; no mechanisms are advanced that would invalidate the reported ratios

28. Professor Walter, in paragraphs 17 to 24, discusses at some length the possible lack of representativeness of those who participated in the study, and seems to imply that a lack of representativeness created "selection biases" that distorted the reported ratios. If this is indeed the intended implication, I point out that selectivity does not necessarily distort the ratios measured in such studies. Walter has not advanced any mechanisms for how it could have

---

[2] Dar-Nimrod I, et al.. Exposure to scientific theories affects women's math performance. *Science*, 2006;314: p 435.

occurred. Indeed, a more complex scenario is required in order to produce such a distortion.

29. Selectivity in and of itself does not necessarily invalidate or 'bias' a comparison. The findings of the classic UK study that compared mortality in smokers and non-smokers were taken to apply to smokers of both sexes, in the general population, despite being based on highly unrepresentative subjects, namely male doctors who did / did not smoke. Naturally, the findings leave open the possibility that the effects would be *somewhat more* or *somewhat less* pronounced in remainder of the general population, who were not included in the study, but it would be very difficult to imagine that the findings would be in the *opposite* direction.

b) Selectivity has to be quite complex in order to produce distortions in ratios in a comparative study

30. Since Walter devoted considerable space to the issue of selection, it is important to understand when selection would, and would not, invalidate the reported ratios. I first use outside examples to illustrate the general phenomenon, and to carefully distinguish between *confounding* bias and *selection* bias.

31. Epidemiologists use the term bias for a systematic distortion, or artefact in the results.

32. *Confounding* (e.g., by age differences in the cited 20-year follow-up study of women) is one source of such artefacts. As noted, if it is caused by a measured variable, it can be corrected for.

33. Another distortion, much more insidious since it is not correctable, is *selection bias.* As described in chapter 11 of the 1980 textbook by Kleinbaum and co-authors, s*election* bias[3] occurs in an epidemiologic study when the *relationship* seen in those who are included in the study is different from what it would be in the target population of interest. In contrast to confounding, it cannot be corrected using the available data on those who are included. Three examples will illustrate the phenomenon of selection bias.

(a) It would be present in an "after the fact" study investigating the relationship between long work hours and adverse pregnancy outcomes, if women medical residents who had already had an adverse pregnancy outcome were more likely to respond than the working wives of male residents who had already had an adverse pregnancy outcome. Residents have been lobbying for many years for shorter work hours. Thus, greater participation of affected residents than affected wives would *exaggerate* the relationship between work hours and pregnancy outcomes.

(b) It would occur if heavier persons, who already had gastric reflux as a result of their higher weight, were less likely, compared to lighter persons who did not have reflux, to participate in a study investigating this association. In this situation, if subjects knew the objective of the study ahead of time, embarrassment on the part of those affected by the relationship might keep them from participating, and thus make the observed relationship weaker than it truly is.

(c) It would occur if, in a study of the relationship between an occupational exposure and ill-health, the researchers only studied

[3] Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research: Principles and Quantitative Methods Wiley, 1982. Other texts (e.g., Rothman KJ Epidemiology: An Introduction, Oxford University Press 2002; Rothman KJ and Greenland S. Modern Epidemiology, Lippincott-Raven, 1998) use the term more generally, to include biases that Kleinbaum describes under confounding.

> (or followed up) those who either were still working in that environment or never worked in it. They would miss those who had worked with the material, but had to retire or change jobs because of the ill-health caused by the exposure. The data from such a study would minimise the relationship, because those affected had been selected out, while those who were immune to its effects would be selectively retained.

34. Thus, the essential element required for *selection bias* to occur is that the probability that a suitable subject is selected-in (or -out) is simultaneously related to *both* the exposure in that subject *and* the outcome or behaviour of interest in the *same* subject. If the probability of inclusion/participation is related to the exposure but not the outcome, or vice versa, the selectivity will not affect odds ratios (chapter 11 of Kleinbaum 1982).

c) *In the examples raised in Walter's discussion of selection bias in the Mintzes study, no such mechanism is postulated as to it would produce distortions in the reported ratios*

35. The above principles imply that, if those in a particular *known* subgroup (e.g. male patients, poorer patients) of those envisioned for a study are less likely to participate than other subgroups (female patients, richer patients), this does not by itself invalidate or bias the odds ratio. Professor Walter does not describe any scenario under which the participation rates would automatically inflate the reported ratios.

36. In order for selection bias to distort, say, the odds ratio for a comparison between Sacramento and Vancouver, we would have to postulate a scenario such as:

> Some of the target Vancouver physicians have practices in which a greater percentage of the patients request DTCA drugs; others have practices in which a smaller percentage of the patients request DTCA drugs; (for personal or practice reasons not related to the measured physician or patient characteristics), the former physicians are less likely to participate than the latter
>
> and, simultaneously
>
> the corresponding participation propensities in the target Sacramento physicians were either reversed, or at a minimum, less related to the amount of DTCA seeking in their practice.

37. None of the objections raised by Walter under the rubric of selection bias postulates such a complex mechanism in respect to the targeted physicians and patients in the two compared cities.

38. If the reasons for participation/non-participation of a patient and physician have nothing to do with whether they would request/prescribe, a low participation rate in both cities, or even a different participation rate in each city, would not, by itself, distort comparisons expressed in terms of ratios.

39. One potential consequence of a low participation rate is that if certain subgroups (e.g. male patients) refused entirely, it would leave open the question as to the generalizability of the results found in females to those in males. But the absence of males would not invalidate the comparison among females.

40. It often occurs that well conducted randomized clinical trials are carried out on only a small fraction of the eligible patients. Again, this non-representativeness does not invalidate the results, or introduce a selection bias. It may limit the generalizability, if the results would be different in those not studied. The proportion of study patients who have a university degree (30% of the Sacramento and 33% of the Vancouver patients) is (I suspect) much higher than it is in the general population of patients. One could make a case that the differences in requesting and prescribing seen in the Mintzes study would be lower if the study had used a less educated sample – if such patients are less likely to ask, or less insistent. I could also envisage the argument that the differences would be greater – if better educated patients are more sceptical about the advertising.

41. For these reasons, the comments in paragraphs 17 to 21 of the Walter affidavit regarding "selection biases" in the Vancouver and Sacramento physicians do not undermine the validity of the ratios reported in the study.

d) No scenario was invoked to show how the methods used to select patients would distort the reported ratios.

42. Professor Walter, in his paragraph 22, was concerned that the mechanisms used to select patients from the participating practices might have induced a selection bias. When I apply the reasoning I outline above, I fail to see how these mechanisms would distort the reported ratios. I presume, for example, that the primary care physicians saw patients mainly on weekdays and thus that

the research assistant carried out the survey on weekdays in both cities. Thus it is not clear to me why there would be day-of-week, or holiday effects, raised by Walter, that could systematically make the observed rate of drug-seeking higher in one city than the other.

43. Moreover, the survey was conducted in March-May in Sacramento (average May temperature 58F) and in June-August in Vancouver (average July temperature 62F). I doubt if this difference would explain much of the difference in drug-seeking, particularly for the DTCA drugs studied. In their interpretation, Mintzes and colleagues do mention the possibility that the survey in Sacramento took place in 2001, some 10 months later than in the 2000 Vancouver survey, and that this may have given the Sacramento patients slightly more exposure, but correctly argue that this would not have affected the within-city comparisons. Moreover, the comparisons did account for the possible confounding effect of socio-economic characteristics of the patients – a factor mentioned by Walter at the end of his paragraph 23.

44. One would need to postulate a more complex selection bias scenario than that in paragraphs 17 to 24 of the Walter affidavit in order for there to be distortion of the exposure - drug-requesting relationship. And the pattern would have to be different in Vancouver and Sacramento in order for it to distort the ratios reported in Table 3.

45. The requirement for selection bias to distort the odds ratios is that the probability that a suitable subject is selected-in (or -out) is simultaneously related to *both* the exposure in that subject *and* the outcome or behaviour of interest in the *same* subject. As is shown in Kleinbaum 1982, if the probability of inclusion/participation is related to the exposure but not the outcome, or vice versa, the selectivity by itself will not affect odds ratios. Even if we had concrete examples of some plausible mechanisms that created non-representativeness of the samples, it would take more than that to produce a distorted estimate of the effect of DTCA on patient and physician behaviours.

3) <u>Planned sample sizes and power</u>

a) Since the study demonstrates a statistically significant "signal," post-study objections about an unrealistically small study size, or about incorrect guesses at the degree of similarity of patient behaviour within physician practices, are not relevant.

46. Paragraph 26 of Professor Walter's affidavit states that Mintzes et al used a "*relatively large* hypothesized impact of DTCA (a 3-fold increase in request rates)", and thus seems to imply that the planned study size was unrealistically small. However, if after a study has been conducted, an appropriate analysis reveals a statistically significant signal, then the magnitude of the "hypothesized impact" used at the planning stage must not have been unrealistically small. If the observed signal were not statistically significant, one could legitimately point to inadequate sample sizes, or unrealistic pre-study assumptions about the magnitude of the hypothesized impact.

47. The Mintzes study found a statistically significant "signal", so a post-study objection to a "relatively large" (and possibly unrealistic) hypothesized effect is not relevant. Moreover, the last line in Walter's paragraph 26 concedes that the hypothesized effect was "borne out."

48. Likewise, paragraph 27 in the Walter affidavit notes that Mintzes et al. did not report how they obtained their pre-study estimates of the degree to which responses in the same practice would be more similar than those in different practices. Again, post-study, in the light of a statistical analysis which used the degree of similarity seen in the actual data, the omission of the method used pre-study, or whether the pre-study estimate was on target, is moot. As above, it would have been more relevant if the authors had underestimated it and the study had then failed to find a statistically significant effect.

49. The analysis in the Mintzes study used a data-based correction for the clustering. Thus, even if the researchers were off in their pre-study estimates of the magnitude of the cluster effect, these estimates did not affect the actual data-analyses. What matters is that the analyses were done correctly, and that the authors corrected for whatever the magnitude of the cluster effect actually was in the data.

4) Differences between Sacramento and Vancouver

a) They are not all as large as is claimed they, and they were adjusted for

50. Paragraphs 28 to 30 of Professor Walter's affidavit highlight the between-city differences in physicians and patients in the Mintzes study and concludes that statistical adjustment will be difficult.

51. I too took immediate note of these differences, which the authors report in their Table 1. Since they were quite aware of these, and made appropriate adjustments for them, and since some of them required relatively little correction, [see 6) below], I am less concerned that they cannot be, or were not, adequately adjusted for.

b) Both the inter-city and intra-city comparisons are helpful

52. Professor Walter (in his paragraph 31) found that the intra-city comparisons are more trustworthy, given what he considers large inter-city differences in the physician and patient and profiles, and given that the reported exposure to DTCA was high even in Vancouver.

53. I find both types of comparisons to be helpful. First, I do not have data on the inter-city differences in the intensity of the advertising, or the amount of information in the advertisements, but I suspect that the intensity is higher, and the information more comprehensive, in Sacramento. If this is so, the differences in drug-requests may reflect this. Even if the intensity and amount of information

per ad is the same, I am struck by the fact that the moderately higher penetration in Sacramento translates into as large an inter-city difference in drug-requesting rates as is observed; I am left wondering what the difference in drug-requesting rates would have been if the difference in the percent penetration in two cities was even more marked.

54. Like Professor Walter, I also like the intra-city comparisons, which may be less affected by any relevant (not just inter-city but) inter-country differences that the authors of the Mintzes study might not have been able to factor out.

5) <u>Validity of Patient data</u>

a) Since, despite random errors in subjects' reports on their exposure to DTCA, the Mintzes study demonstrated a statistically significant "signal," it is likely that an even larger signal would have been seen in the absence of such errors.

55. Paragraph 32 of Professor Walter's affidavit raises the importance of accurate patient responses about exposure to DTCA, and the desirability of validating them. However, the observed statistically significant differences probably represent underestimates.

56. The more accurate the responses, and the less the "noise," the more chance there is to detect differences in drug-seeking between groups of patients who have seen different amounts of advertising. One would have to assume that there are *some* errors in the self-reports. Despite this, statistically

significant differences in drug-seeking were found. Had the researchers been able to correct for any random errors in self-report, the differences would have been even larger; in other words, the gradients seen in the study are probably underestimates relative to what they would have been if responses had been more accurate.

b) No scenario was provided in the Walter affidavit as to how a more complex pattern of errors in reporting of DTCA exposure could, by artefact alone, create the ratios observed.

57. For inaccurate responses to have *exaggerated* the true differences in drug-seeking, and made the gradient in Figure 2 higher than it is in reality, one would have to postulate a mechanism, such as those who saw 2 advertised products reported seeing 1 and those who saw 3 reported seeing 4, etc. I cannot advance a plausible reason why this would have happened.

c) No scenario was provided in the Walter affidavit as to how a more complex pattern of errors in reporting their drug-seeking behaviour could, by artefact alone, create the ratios observed.

58. Had the study failed to show any differences in drug-seeking behaviour, it would be appropriate to attribute the failure in part to less than perfect self-reports of the behaviour. However, if, despite less than perfect recall, significant differences are found, this means that the differences would have been stronger if the researchers had less fallible data on what actually took place.

59. Moreover, errors in self-report are likely to apply equally in Vancouver and in Sacramento. We do not have plausible reasons to believe that patients in Sacramento are less forgetful, or over-report more, than those in Vancouver.

6) Issues in the statistical analysis

a) The sample size in the Mintzes study was adequate to support the needed statistical adjustments

60. Paragraph 36 of Professor Walter's affidavit claims that two sample size constraints, the number of factors to be adjusted for, and the fact that the dataset for Tables 3 and 5 contains only 74 instances where the patient requested a DTCA drug, make the sample size too small to reliably support an analysis of this complexity. In support of this assertion, the paragraph cited a rule of thumb that there be 10 such observations for every term in the adjustment equation (Professor Walter counted at least 12 terms in the equation). As I explain below, applied to the data in this study, and to the role of these factors in this study, this assertion is a considerable overstatement.

61. The purpose of the types of statistical adjustments referred to is to (mathematically) correct for the (unavoidable) differences between the compared groups with respect to factors that affect the behaviours of interest (e.g., drug-seeking) that would otherwise distort or confuse the comparison. First, by way of an expository example, suppose we wished to determine how much faster a group of persons, who are given a certain encouragement or other potential

performance enhancing agent, can run a mile than a group who were not. Suppose we were limited to an "observational" (i.e. non-experimental) study in which the two groups were not formed by randomization. We would want to make the comparison fairer by adjusting (or handicapping) for the between-group differences in factors that affect the time it takes to run a mile, such as age, fitness, gender, etc. Thus we need values for the adjustment factors (such as the number of additional seconds/minutes to be allowed for every year one group is on average older than another, or for every 10% more in the one group who are women). Since the appropriate magnitudes of such correction factors are seldom available externally, the values are typically derived from the same dataset in which the comparison of interest is being evaluated. One does pay a statistical price, in the form of a larger margin of error around the reported odds ratios, for having to derive them from the dataset.

62. Contrary to the impression that may have been conveyed by paragraph 15 of Professor Walter's affidavit about the price (in terms of larger samples sizes) needed for statistical adjustment, the 95% confidence intervals (CIs) reported by Mintzes and colleagues already include the "price" of having to adjust for confounding variables.

63. Two factors affect the stability of the resulting adjusted ratios. The first is the degree to which the groups were different with respect to the measured factors that affect the target behaviour. Of the 8 adjustment variables discussed in paragraph 36 in Professor Walter's affidavit, the profiles for three of

the factors in the Mintzes study (age, sex, age and general health status) were virtually identical in the Sacramento and Vancouver patients: {average 50y versus 48y}, {36.5% male versus 33.2% male} and {82.9% in good-excellent health versus 83.8%}. Thus, in Table 2 of that study, the comparisons require minimal correction for these three factors (even if these three factors were to have a large influence on drug-seeking behaviour).

64. The *education* profiles were somewhat different, but I am not well enough versed in patient drug-seeking behaviour research to know, if education were the only relevant factor that influenced behaviour, and if we could match the groups on all other relevant factors, which group would have the greater drug-seeking tendency as a result of this difference in their education profiles. However, since the profile differences are not large, and since the comparison adjusted for them, I am not concerned that they distort the comparison.

65. The largest differences were in the patient's *income* and in *who pays* for medications; the study authors did adjust for these.

66. There were also some differences in the age and gender profiles of the patients' physicians; these were also corrected for.

67. Judging from the generally small differences between the unadjusted and adjusted Odds Ratios (ORs) in Tables 3 and 5, the effects of patient's income and in who pays for medications must either have been small,

or else had effects that, when applied to the differences in the profiles, cancelled out in the correction. This would be akin to a situation where we were comparing the times to run a mile in a group who received a potential performance-enhancing intervention who were younger than, but contained a greater proportion of females, than the comparison group who did not receive such motivation.

68. Thus, in summary, whereas the Mintzes study considered eight factors as possible correction factors in Tables 3 and 5, one can see directly from Table 1 that little correction was required for three of them, and from the tables themselves that the corrections for the others did not substantially affect the ratios.

b) The "criterion of 10:1" is conservative when the purpose is adjustment, and most of the inter-related factors were used merely for adjustment

69. Paragraph 37 in Professor Walter's affidavit expressed particular concern about the 10:1 criterion in Table 4 "when the factors included in the model may themselves be related to one another." It states that work by others has shown that conclusions from such models fitted with insufficient sample size can be substantially in error with respect to the magnitude, precision, statistical significance, and *even the direction* of the associations indicated in the results. Applied to this particular study, these statements are considerable overstatements.

70. The 10 for 1 criterion is a rough rule of thumb, and is sometimes violated by reputable and highly credible statisticians. The formal research on this topic, and in particular the implications for correction for confounding, is quite limited.

71. However, a recent article[4], entitled "Relaxing the Rule of Ten Events per Variable (10EPV) in Logistic and Cox Regression", is highly relevant to the statistical analyses in the Mintzes study, and to the issues raised by Professor Walter. In this article, the authors, based on simulations, conclude that the rule of thumb of a minimum of 10 outcome events per predictor variable (EPV) may be too conservative. They found a range of circumstances in which coverage and bias were within acceptable levels despite less than 10 EPV, as well as other factors that were as influential as or more influential than EPV. They concluded that this rule can be relaxed, in particular for sensitivity analyses undertaken to demonstrate adequate control of confounding. Further, in their discussion, they state that when a statistically significant association is found in a model with 5–9 EPV, only a minor degree of extra caution is warranted, in particular for plausible and highly significant associations hypothesized *a priori.*

72. While the statements about 10EPV made by Professor Walter are more relevant for certain types of studies, they are less pertinent in this type of study. They are most relevant when one is interested in *estimating the*

independent effect of each factor. For example, suppose we were interested in how much each of the factors waist size, BMI, collar size, belt size, *separately* affects people's times to run a mile. Thus, we would want to know how much longer it takes for persons with collar size x+1 than those with collar size x, but who all have the same waist size, BMI, and chest size. It is in this context, with *highly interrelated variables such as those just listed*, that one can obtain the (cited) "unreliable" estimates of the *unique* contribution of each factor.

73. The concerns are less relevant when (as in the simulations carried out by Vittinghoff and McCulloch, cited above) the adjustment factors are not of primary interest, but are included in the adjustment model for the express purpose of correcting the primary comparison (enhanced vs. not in the hypothetical example; more exposed versus less exposed to DTCA in the Mintzes study) for the differences in the profiles of the compared groups. In our example, when the primary goal is to isolate the one effect of interest (enhancement), we are not concerned whether the neck size is more important than chest size or waist size than BMI. We adjust for the set of such confounding variables, or some combination of them.[5]

---

[4] E Vittinghoff and CE McCulloch. "Relaxing the Rule of Ten Events per Variable (10EPV) in Logistic and Cox Regression". American Journal of Epidemiology (2007), Vol. 165, No. 6, pp 710–718.

[5] The same considerations apply when a financial institution uses factors collected on previous applicants who applied for a loan to develop a *prediction* equation for the probability that a new applicant will not default on his/hers. The fact that some predictor variables may themselves be related to one another is not a serious concern, since the purpose is not to find out exactly how much each item *independently* influences the probability, but rather *collectively* how informative they are. Including two closely related variables, where the directions of the associations indicated in the results might even be opposite from what one would expect, has little impact on the quality of the predictions. What matters more in the prediction context is how large was the pool of candidate factors in relation to the numbers of defaulters in the dataset being used to construct the prediction equation.

74. Paragraph 36 in Professor Walter's affidavit makes specific reference to the four explanatory variables of interest shown in the body of Table 4. However, based on the widths of the reported confidence intervals, the four explanatory variables must not have been that highly correlated. The extent to which these variables are related to each other, and to the probability of requesting a DTCA drug, is reflected in the Odds Ratios (ORs), and especially in the margins of error. Had they all been tightly related to each other (e.g. if those who had seen advertisements for > 3 drugs also tended to be the same patients who had a condition treated by one of an advertised drug, and used advertising as an information source, and lived in Sacramento), they would yield ORs with much larger margins of error than appear in the table.

75. In summary, I believe that the estimated ORs for the four explanatory variables of interest in Table 4, and for the factors of interest in Tables 3 and 5, are unlikely to be substantially in error with respect to the magnitude, precision, statistical significance, and the direction of the associations indicated in the results. The widths of the confidence intervals (CIs) in Table 4 indicate that the factors were not so highly related as to preclude estimating their separate effects. Some of the differences in Table 1 were minor, and adjustments for them in Table 3 and 5 did not substantially change the unadjusted ORs shown one column to the left. These conclusions are in accord with the findings of Vittinghoff concerning inferences from models with as few as 5–9 EPV.

c) Despite some reservations on my and Professor Walter's part as to the data quality, it appears that physicians are equally ambivalent about prescriptions for DTCA and non-DTCA drugs that are written in response to patient requests.

76. Paragraphs 41 to 43 of Professor Walter's affidavit deal with the way in which physicians' opinions about the "appropriateness" of the prescribed drugs were measured and the limitations of the 3-point scale (the likelihood that they would prescribe the same drug to other similar patients). I note that Mintzes and colleagues specifically state (in the second paragraph in the Interpretation section on page 411) that they (the authors) "could not evaluate treatment appropriateness".

77. Table 6 in the Mintzes study also shows physicians' opinions regarding *non*-DTCA drugs that were requested by patients. They seemed to be fairly similar to those about DTCA drugs requested by patients, at least in Sacramento. I will use the "likely to prescribe for other similar patients" category, the complement of that used by Mintzes et al., and separate the requested drugs into DTCA and not DTCA (Mintzes el at used DTCA drug, and "any drug" (i.e. "DTCA plus all others"). The physicians rated almost 90% of the prescriptions that were *not* requested by patients to be in the "less ambivalent" category (i.e. they "would have been a very likely choice for other similar patients"), but only approximately 55% of those requested by patients to be in this same category. Although this 55% was considerably lower than the 90%, it was equally low for prescriptions of *both DTCA and non-DTCA* drugs in Sacramento (52% vs. 55% respectively, based on denominators of 42 and 56). In Vancouver the

percentages were 44% versus 74% respectively, but based on much smaller denominators (18 and 27). These percentages could to be taken to mean that physicians' opinions regarding many of the prescriptions they issue in response to patient requests are in the "ambivalent" category – regardless of the source (DTCA or other) of the patient's awareness about the drug.

78. The authors of the Mintzes study had to be careful and non-directive in their assessment of physician's opinions about patient-requested drugs, and one can understand why it was difficult to use more exact methods in this sensitive and hurried context. Another way to assess a physician's opinion regarding appropriateness or not of requested DTCA drugs might be via a survey which used only hypothetical patients, or to assess it indirectly in an experimental setting where a physician does not having to justify it after having prescribed it to a particular patient.

7) Interpretation of Results

a) Does DTCA lead to more requests for prescriptions of these DTCA drugs?

79. The study found that the request rate was higher in those patients exposed to more DTCA. When a well done observational study finds a higher rate in those more heavily exposed, commentators will nevertheless search for overlooked non-causal alternative explanations. But in this context, we should consider the obvious rather than the overlooked one: the stated message in much of DTCA is "ask your doctor if this drug is right for you." Unless I would

have reason to think that advertisers do not wish to create more requests, or that their advertising is not successful, I can only reason that the answer to the question I posed above is yes. In paragraphs 46 and 47 of his affidavit, Professor Walter, despite the concerns raised in his earlier sections as to the validity of the reported request-ratios, seems, like I do, to accept the obvious.

b) Does DTCA lead to more prescriptions of DTCA drugs?

80. The study also found that the prescription rate (not just the request rate) was higher in those patients exposed to more DTCA. Again, in light of the intent of DTCA, and despite any statistical uncertainties arising from the fact that we cannot study this experimentally, and from the amount of data we have, it makes sense to interpret it as causal.

c) Does DTCA lead to more prescriptions of DTCA drugs about which physicians are ambivalent without also increasing the prescribing of DTCA drugs about which they are ambivalent?

81. Both Professor Walter and I have some concerns about the quality and interpretation of the data in Table 6. However, despite my concerns about these, the study data (and other studies) indicate physicians appear to be ambivalent about some of the prescriptions they write in response to patient requests. In Sacramento, this was equally the case for both DTCA and non-DTCA drugs, whereas in Vancouver it seemed to be more the case for the DTCA ones. Even with this lack of precision as to the differential effects, I conclude, based on the increased requests and prescribing generated by DTCA, and the

reported ambivalence in a fraction of these, that DTCA will increase both the prescribing of DTCA drugs about which physicians are not ambivalent, and the prescribing of DTCA drugs about which they are ambivalent.

D. SUMMARY AND OPINION

82. In my opinion, the objections concerning (1) selection bias; (2) sample size; (3) inaccuracies in patient responses; and (4) statistical adjustment raised in the affidavit of Stephen Walter should be given less weight than he attributes to them. (1) No concrete or plausible selection scenarios that would lead to distorted odds ratios were suggested in his affidavit; (2) the sample size did not need to be as large as he argued for, since several factors were already quite similar in the compared cities, and the sample size criterion he invoked has been shown to be conservative in this type of data-analysis; (3) errors in patient reporting would most likely have attenuated the differences, not inflated them; the fact that statistically significant differences were observed despite these alleged inadequacies in sample size and data quality suggests that the signal would have been even stronger had the sample sizes been larger and the errors smaller; and (4) the authors of the study did carry out appropriate statistical adjustments.

83. Thus, since the objections raised cannot be substantiated, I conclude that the Mintzes study does provide evidence about the effects of DTCA. Moreover, the differences in drug-requesting and drug prescribing observed in the Mintzes study have additional plausibility, since they are in line with what one would expect from successful DTCA.

84. I make this affidavit in response to CanWest's application, and for no other or improper purpose.

AFFIRMED before me at the City of Montreal, in the Province of Quebec, this 28th day of November, 2007.

Michelle Dion
Commissioner for Taking Affidavits

James Hanley
JAMES HANLEY