Imperial College, University of London, together with the Associateship of the Royal College of Science.

3. I am a Fellow of the Royal Statistical Society, a Fellow of the American Statistical Association, and a member of other professional societies and associations. I have authored or co-authored over 300 refereed journal articles and book chapters and over 300 other communications, including book reviews, editorials, letters, abstracts, presentations, and other publications.

4. The impact of scientific research can be assessed by how often published research reports are referenced by other scientists. The Institute for Scientific Information (ISI) maintains a "*Web of Science*" web-based service that can be used to calculate the world-wide citation rates for papers published in journals that it indexes. Such an analysis was last done for my work about 2 years ago, based on publications on which I was the sole author or a co-author. The results of that analysis showed that at that time there had been a cumulative total of almost 10,000 citations to my work. This analysis was based on 247 of my papers that had appeared in journals that are covered by the ISI. At that time, I had also published 29 papers in other journals not covered by ISI.

5. As well as evaluating the impact of individual research papers, the ISI also calculates the overall Impact Factor for each journal, by assessing the average citation rates for all of the papers it publishes. These calculations are done by categories of journals, because scientific citation patterns differ considerably between scientific

disciplines. A journal's Impact Factor is generally recognized as a measure of the journal's quality, and it is very desirable to scientific researchers to publish their work in journals with high possible Impact Factors. Peer reviewing in these journals is correspondingly very stringent, and the acceptance rate among all the papers they receive for consideration is relatively low. In recent years, I have authored at least one paper in each of the 8 leading General and Internal Medicine journals (there are 102 journals in this category), having the highest Impact Factors. My work has also frequently appeared in journals that publish biostatistical methodology, that are also very highly rated by ISI in their category.

6. I work in the fields of epidemiology and biostatistics. My own research interests in epidemiology have included studies of cancer, disease screening, and environmental health. My interests in biostatistical methods include measurement error; diagnostic and screening test evaluation; observer agreement; risk assessment and communication; and spatial and temporal data. I also collaborate frequently with clinicians on the design and analysis of medical research studies, including randomised clinical trials of new forms of therapy, and observational studies of clinical practice and decision making.

7. I was an Editor of the *American Journal of Epidemiology* for 12 years (1988-99) – this is the leading journal in its class, as assessed by ISI. I am also the Section Editor for Clincal Epidemiology in the Wiley *Encyclopedia of Biostatistics*. I have served as the Chair of Biostatistics in the International Clinical Epidemiology Network, and I have been involved with the development of clinical epidemiology in various countries in Asia,

Latin America and Africa. I teach in the graduate school of McMaster University, and I am a past coordinator of its educational programs in Clinical Epidemiology.

8. My work has been recognised nationally and internationally, and I have won several professional honors and awards, as shown in detail in my curriculum vitae. I have also been asked to serve on a variety of expert panels for organisations such as the Canadian Institutes of Health Research, Health Canada, the United States National Institutes of Health, and the World Health Organisation. My curriculum vitae is attached as **Exhibit "A"**.

Scope of this Affidavit

9. I have been retained by counsel for CanWest Mediaworks Inc. ("CanWest") to provide an opinion on the methods and results in the study by Mintzes and colleagues that is published in the Canadian Medical Association Journal, 2003, entitled "How does direct-to-consumer advertising (DTCA) affect prescribing? A survey in primary care environments with and without DTCA" (the "Mintzes Study").. This study is relied on by various affiants for the Attorney General of Canada in this case on the possible effects of direct-to-consumer advertising (DTCA). A copy of the Mintzes Study is attached as **Exhibit "B"**.

10. DTCA can be characterized as an "exposure" that may influence outcomes such as patients' behaviour, in terms of their seeking care from their primary health care physicians. Patient behaviour – in particular requesting certain prescriptions associated

with DTCA – may in turn influence physicians in how they prescribe for their patients. The Mintzes Study addresses both these questions by studying primary care physicians and their patients in two populations in the areas of Vancouver (BC) and Sacramento (California).

Methodologic challenges in evaluating the effects of DTCA

1) Observational nature of the data

11. The ideal study design to investigate the relationship between an exposure and outcomes of interest employs a randomized assignment of individuals in the study population to be "exposed" or not. Randomisation provides assurance that the exposed and unexposed participants in the study are equal, on average, apart from their exposure status. This has the important advantage of avoiding bias that might be otherwise be associated with other determinants of the study outcomes. Depending on specific study circumstances, these other determinants may be known and measured (in which case statistical adjustment can be attempted), known but unmeasured (typically because of reasons to do with infeasibility), or unknown (and therefore unmeasured). Such determinants are known as "confounders" of the effect of exposure of interest. The associated bias is commonly referred to as "confounding".

12. While a randomized design protects against bias from confounders, no such protection exists in observational studies where randomized assignment of exposure is not used. The assessment of evidence concerning a particular exposure-outcome

association from an observational study must therefore critically examine the possibility of confounding as an alternative explanation of the study findings. Because of the nature of DTCA, the Mintzes Study is observational, and hence the possibility of confounding of the effect of DTCA exposure on patient and physician outcomes must be carefully considered.

13. Health-care seeking behaviour of patients and prescribing habits of physicians are complex matters, and each will have numerous potential determinants. Such determinants would include factors such as the patient's age, sex, health status, income, education, method of drug payment, and the physician's sex and graduation year. (These particular factors were measured in the Mintzes Study and used in the analysis). Other potential determinants, that were measured in the Mintzes Study but not used in the analysis, include the number of patients in the physician's practice and the length of the patient-physician relationship. Finally, there may well be other determinants apparently not captured in the Mintzes Study, and would include but not be limited to factors such as details of the physician's training, and patient family composition and size, and ethnicity.

14. The goal of the Mintzes Study is to isolate the effect of DTCA exposure on patient and physician outcomes, free of the confounding effects of these other determinants. Because of their large number and the likely strength of their own associations with patient and physician outcomes, these determinants collectively make it very difficult to isolate the specific effect of DTCA. In the framework of an observational study, such as Mintzes, there is therefore strong potential for confounding

by the other determinants. The confounding problem pertains to both inter-city comparisons (Sacramento vs. Vancouver) and intra-city comparisons of individuals.

15. To attempt to isolate the specific effect of DTCA, one must rely on complex multivariate statistical adjustment of the data, to take account of the many other differences that exist between the groups of individuals who are exposed or not exposed to DTCA. Such statistical analyses typically require large samples when numerous confounders are involved, such as in the Mintzes Study, especially when the groups of subjects being compared differ substantially. As discussed later in more detail, the Mintzes Study demonstrates numerous substantial differences between the Sacramento and Vancouver populations, implying that that isolation of the particular effect of DTCA between those two communities will be particularly challenging. In my opinion, the relevant sample sizes in the Mintzes Study are not sufficient to reliably support the complex statistical adjustments needed for the many likely confounders of the DTCA effect.

16. The same problem also applies to the intra-city comparisons reported by Mintzes. Additionally, the statistical adjustments applied to both inter-city and intra-city comparisons are by necessity limited to only confounders that have been recognized and measured in the study, but they cannot take account of unrecognized or unmeasured confounders.

2) Selection biases: general considerations

17. A further issue in evaluating the evidence from a study such as that by Mintzes is the possibility of selection bias in identifying the participants. In a high quality study, one would first define the target population of participants, and then attempt to determine if the study participants constitute a representative sample of that population. In the case of the Mintzes Study, we need to ask if the patients were representative of all patients in the corresponding populations, and if the physicians were representative of all doctors serving those populations. High study quality in this respect would be indicated by clear specification of the target populations, and a high participation rate among the individuals who are sampled from those populations, and who are then approached and asked to take part in the study. Furthermore, and even if response rates are generally high, it is still very desirable to investigate if the study participants are representative, through comparisons between the participants and non-participants with respect to observable characteristics.

2.1) Selection biases – Vancouver physicians.

18. The target population of physicians in the Vancouver portion of the Mintzes Study was well defined, by using the Medical Directory for the area, and a random sample of 200 doctors was drawn from that Directory. However, only 78% of the sample doctors were actually contacted by phone, and only 103 met the study inclusion criteria. Of these 103 physicians, only 23 agreed to participate. Overall this yields a response rate of 23/200, or about 11% - a figure that would usually be regarded as extremely low for

an epidemiological study. The low response rate increases the potential for selection bias to affect and bias the results. Given this low response rate, one would therefore particularly expect to see some investigation of the representativeness of the participants to the target population, but no such investigation is reported by Mintzes. As a result, while the potential for important selection biases exist, one cannot actually examine the question with the available evidence.

19. In addition to the 23 Vancouver physicians who complied initially, the study also included a further 17 doctors who were practice partners of the 23. While these additional doctors enhance the sample size, in my opinion it is problematic that the extra 17 were not (apparently) part of the original sample of 200. Opportunistic inclusion of additional participants in this way violates the original sampling design. Furthermore, there is again no analysis of the likely representativeness or otherwise of this supplemental group of participants. Here in particular there is a distinct possibility of selection bias, because the 17 supplemental participants were by definition all working in group practices (as opposed to solo practices, which could not be accessed by this route), and they were associated with the 23 physicians who had already agreed to take part.

2.2) Selection biases – Sacramento physicians.

20. In Sacramento, 62 primary care physicians working in the University of California, Davis network were invited to take part. Mintzes et al. do not indicate if this is the

number of *all* the relevant physicians or not, so the question of sampling representativeness may or may not be an issue here.

21. Among the 62 physicians who were approached, 38 (61%) agreed to take part. While this response rate is better in Vancouver, it is still quite low, and low enough that selection bias could again be a factor influencing the final results. The Mintzes report provides no information about characteristics of the Sacramento participants vs. non-participants, and hence we cannot assess their representativeness.

2.3) Selection biases – patients

22. A similar concern about selection bias exists with respect to the recruitment of patients in both Sacramento and Vancouver. According to Mintzes et al., days for patient recruitment were pre-selected, but the way that pre-selection was done is not specified. It is therefore unclear if days were chosen at random, and what attempts (if any) were made to deal with day-of-week, holiday, or seasonal effects. All of these effects could be related to the type of patient attending a primary care clinic, with respect to demographics and urgency of the case. These case features could, in turn, be related to DTCA exposure, and thus bias the results.

23. The response rate of patients on the selected days is reported as 61% (for Vancouver and Sacramento combined). Again we have no information on the non-participant patients, and so sample representativeness cannot be assessed. (In fact, additionally, although Mintzes et al. report that a research assistant contacted

consecutive patients, it is not completely clear that this method targeted *all* patients on the selected days, as would be the most valid method). Relationships could exist between the willingness of patients to participate and factors such as their time availability and socio-demographic status, and those factors could in turn be related to the likelihood of having been exposed to DTCA. If so, the results would be confounded, and their interpretation consequently problematic.

2.4) Selection biases – summary

24. Because of the lack of information on non-participating physicians and patients, the potential for non-representativeness of the participants in the Mintzes Study cannot be assessed. However plausible mechanisms certainly exist by which one can imagine non-representativeness of the samples would have occurred, and occurred in ways that are related to the likelihood of DTCA exposure on the part of patients, or to prescribing patterns on the part of physicians. If these concerns are valid, the result will be biased conclusions on the effect of DTCA on patient and physicians outcomes. In the Mintzes Study in particular, the low response rates for both physicians and patients heightens this area of concern about the results.

3) Method of obtaining patient consent to participate

25. Patients were contacted while in the doctor's waiting room, to obtain consent to be in the study. The specific information provided to patients at that stage is not detailed by Mintzes, but one would prefer not to heighten awareness of the study objectives in

any way. Specifically, if the information conveyed by the informed consent process mentioned anything about prescriptions, then the recruitment process itself might have influenced subsequent behaviour. For instance, mention of the pattern of prescription use might itself lead to increased request rates by patients.

4) Planned sample sizes and study power

26. Mintzes et al. report that the required sample size of paired patient-physician data (636) would be sufficient to have 80% statistical power to detect effects such as a 2% vs. 6% request rate for DTCA drugs in exposed vs. unexposed patients respectively. This effect size represents a relatively large hypothesized impact of DTCA (a 3-fold increase in request rates), but that is borne out by the subsequent results.

27. The sample size calculation appropriately adjusts for the effect of sampling patients who are clustered within the same physician practice, although the details of how the magnitude of the cluster effect was estimated are lacking. However, there is apparently no adjustment for possible cluster effects associated with sampling doctors from the same group practices. As noted earlier, we can ascertain from the Mintzes report that a high percentage of the Vancouver physicians were indeed working in group practices. The situation in the Sacramento sample of physicians remains unknown, although one might speculate that many of them, by being associated with a university health care system, were also in group practices. To the extent that prescribing habits might be similar among members of the same group practice, failure to include this

effect in the analysis would lead to over-optimism about the precision of the statistical findings, and corresponding over-statement of their statistical significance.

5) Differences between Sacramento and Vancouver

28. As shown in Table 1 of the Mintzes Study, many characteristics of the Vancouver and Sacramento samples of physicians and their patients were substantially different. For example, relative to Vancouver, physicians in Sacramento were more often male, less often in part-time practice, and were more often salaried. Also, although all participating physicians were in primary care, those in Sacramento were a mixture of general internists and family medicine, while those in Vancouver were all family physicians.

29. Relative to Vancouver, patients in Sacramento had higher incomes, higher educational levels, paid more often for the partial costs of medicines, and were more often studied on a first appointment with the physician.

30. Inter-city differences in physician and patient characteristics can be a good thing in the sense of potentially increasing the generalisability of the study findings, but they unfortunately add to the difficulty of making valid comparisons *between* Vancouver and Sacramento, particularly with respect to identifying specific effects of DTCA between those two populations. Between-city comparisons will require the use of statistical models to adjust for the several or many factors that might differ between the cities, and hence act as potential confounders of the DTCA effect on outcomes. Given the strong

differences between Sacramento and Vancouver that were identified by Mintzes et al., one must conclude that statistical adjustment will be difficult, especially with the limited sample sizes available.

31. With the above thought in mind, and given the strong differences between the cities that were identified in Table 1, I found the results based on inter-city comparisons to be less trustworthy than the comparisons made *within* cities, evaluating differences between patients exposed or unexposed to DTCA. To some extent, the same was possibly true in the minds of the investigators; when very high exposure rates to DTCA were found in both cities, they report many of their findings between patients in the same city, separately for Vancouver and Sacramento.

6) Validity of patient data

32. Much of the analysis in Mintzes et al. is based on the patient self-report of having seen advertising for various and specific prescriptions. Given the central importance of these data, some validation of patient responses about exposure to DTCA and other advertising would have been desirable. The authors are critical of other studies in this area which have relied on recall of past behaviours, but the same objection surely applies to their own study participants' ability to accurately recall specific advertising that they may or may not have seen.

33. A similar concern exists with respect to patient recall about the physician encounter. Other studies have shown less than perfect recall by patients of health care

advice they have received, even if only a few minutes earlier in a medical appointment. Recall of requests for prescriptions that they themselves have initiated might arguably be more reliably reported, but it would have been interesting to validate that assumption, for example by comparing patient reports of prescription requests with the corresponding data from their physicians.

7) Issues in the statistical analysis

34. While I am concerned about the representativeness and quality of the data in the Mintzes Study, there are some additional points of concern in the analysis, that may or may not influence the results, as outlined below.

7.1) Adjustment for confounders

35. Many of the analyses rely on extensive statistical adjustments in order to try to take care of the numerous confounding variables, in an attempt to isolate the DTCA effect from the other determinants of physician and patient outcomes. Examples are found in Mintzes' Tables 2, 3, 4, 5 and 6, whose footnotes list the covariates that are adjusted for in each analysis, typically a multivariate GEE model.

36. One common criterion for the validity of such statistical models is that there be a minimum of at least 10 observations (or, for analyses with a binary outcome, 10 outcome events) per model parameter. Each covariate requires at least one parameter, and more in the case of a factor such as patient educational level, which (according to

Table 1) has three levels. The model shown in Table 4 has a sample size of 74 (patients who reported requesting a DTCA drug), and there are 8 confounders adjusted for, implying at least 8 parameters – possibly more, depending on exactly how they were incorporated in the model. There are also four explanatory factors of interest shown in Table 4, each involving one parameter. Collectively we therefore have at least 12 main parameters being estimated from 74 observations, and there may be additional parameters implied by the clustered sampling structure that is incorporated in the GEE methodology. The sample size is therefore clearly too small to support an analysis of this complexity with any reliability.

37. 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. These concerns are particularly pertinent when the factors included in the model may themselves be related to one another. Examples of this related to Mintzes' Table 4 include the likely relationships between patient education and income, and between the various indicators of DTCA exposure ("has seen advertisements for > 3 listed drugs", "has a condition treated by an advertised drug", and "uses advertising as an information source").

38. Similar concerns pertain to the findings given in other tables of results in Mintzes et al., although some of them do enjoy larger sample sizes, particularly when it is patient data being analysed. In my opinion, the findings of all the complex models being reported must be interpreted cautiously.

7.2) Missing data

39. There were some missing data for confounders (such as income). Imputation was used for the missing values, using regressions based on known covariates that are correlated with the missing variable. While imputation is a common technique to deal with missingness, Mintzes et al. do not discuss the sensitivity of their conclusions to the imputed values, so the potential for bias to have been introduced by the imputation process cannot be dismissed.

7.3) Possible numerical error

40. As a smaller point, there may be a possible numerical error in the adjusted odds ratio reported in Mintzes (page 409) associated with prescribing advertised drugs. The adjusted odds ratio is given as 2.2, based on prescribing rates in Sacramento and Vancouver of 77.6% and 72.0% respectively. Although the odds ratio here is adjusted for other factors, its value of 2.2 still seems large in relation to the underlying percentages on which it is based, and whose difference is small.

7.4) Interpretation of physician outcomes

41. The Mintzes results include the finding that about 50% of prescriptions for requested DTCA drugs were judged by physicians as a "possible" or "unlikely" choice for a similar patient. This is a difficult and perhaps surprising result to interpret. One

possible interpretation is that physicians are somehow being pressured into undesirable prescribing by their patients. On the other hand, physicians may be saying here that patients are all "different" (despite having being asked about "similar" patients), and that other features of individual patients would prevail in their decision making.

42. Interpretation of this particular result would be enhanced by knowing the breakdown of the "possible" and "unlikely" responses into separate categories. "Possible" was the middle of three categories offered, and as such (I presume) was intended to represent a neutral stance. Response patterns to questions of this type can be strongly dependent on the exact wording and the context of that wording. For what it might be worth, my personal interpretation of the word "possible" is of leaning towards an actual re-use of the drug for a similar patient, rather than strictly neutral. If some study participants thought the same way, one would conclude that the middle category ("possible") is not really neutral. Some other type of scale (such as a visual analogue) might have done better.

43. A separate breakdown of the "unlikely" category would be very informative. For instance, if the "unlikely" responses represented most of the combined "possible" and "unlikely" responses, one would probably conclude that patient pressure to prescribe was important in the clinical decision making. On the other hand, if "unlikely" responses were rare, that explanation would be relatively untenable.

8) Interpretation of study results

44. Mintzes et al. themselves present two main interpretations of their study data; first that

(I) "more advertising leads to more requests for advertised medicines",

and

(II) that it leads to "more prescriptions".

(cf. Mintzes paper, "Interpretation" section of the Abstract). However, the methodological shortcomings of the study that I have identified admit the possibility of various other interpretations of its findings.

45. Concerning the first of Mintzes' conclusions, the data showing an association between patients requesting a DTCA drug and having been exposed to DTCA could be at least partly because concerned patients (e.g. those with a relevant condition) took it upon themselves to get drug information before consulting the doctor. So rather than having a steering effect on patients, one could argue that advertising simply provides relevant information that patients were looking for anyway.

46. Concerning the second conclusion (II), even accepting that there is a higher rate of prescriptions by their physicians among patients who request them, this finding might also reflect entirely appropriate clinical decision making.

47. For both of these outcomes, one cannot necessarily conclude that advertising is the causal agent, in the sense of "causing" more requests or prescriptions that would not otherwise have occurred. And even if the findings were regarded as causal, they might be interpreted as desirable or undesirable, depending on one's point of view.

48. Other commentators have discussed the question of whether changes in prescribing by physicians as a result of patient requests were or were not appropriate. In an exchange between Dr. Mintzes and Dr. Temple at a public hearing on DTCA held by the U.S. Food and Drugs Administration ("FDA") in September, 2003, Dr. Temple (listed as Director, Office of Medical Policy, Center for Drug Evaluation and Research at the FDA) interprets the findings to suggest that such changes represented "...no implied inappropriateness", and later that "My presumption is they [*the study doctors*] are not telling you I gave the person a drug the person didn't need" (my italics added). In other words, even if we accept that some prescribing changes have occurred as a result of patient requests, we cannot necessarily conclude that those changes were positive or negative with respect to appropriateness to the patients in question. An extract from the transcript of the FDA hearing of September 2, 2003 is attached as **Exhibit "C"**.

49. In later questioning at the FDA hearing, Dr. Temple suggests that more detail would have been desirable in the Mintzes Study, and that a future study should allow for more explanation by the physician for prescribing actions made as a result of patient requests. I agree with these comments.

50. In my opinion, the Mintzes Study does not provide any evidence on whether prescribing changes in response to patient requests were appropriate or not. Investigation of this question could be carried out by a clinical audit. This technique, which is commonly used to evaluate outcomes in medical research studies and in assessments of clinical practice patterns, involves an independent review of all the documents pertaining to a set of clinical encounters. Using pre-defined criteria, independent reviewers can classify what proportion of physicians' actions (such as prescribing particular medications) were appropriate, given the clinical presentation of the patient. Lacking such an audit for the Mintzes patient samples, one cannot comment on the appropriateness or otherwise of the physician behaviour in this study.

Conclusions

51. While I find the Mintzes Study to have produced some potentially interesting results, I am concerned about their validity. As outlined above in detail, major methodological limitations of the Mintzes Study pertain with respect to:

- The observational nature of the data, and the strong potential for confounding
- Selection biases that may have affected recruitment of physicians and patients into the study. These biases may have led to further confounding, and certainly limit the generalisability of the results.
- Numerous strong observed differences between Sacramento and Vancouver, severely limiting the potential to isolate the effects of DTCA through comparisons of those cities.

- Limited sample sizes in the study, that threaten the validity of the complex statistical adjustments needed for the numerous confounders that were identified.
- Questionable validity of patient self-reports of exposure to DTCA.
- Lack of evidence concerning the appropriateness or otherwise of clinical actions that may have been taken as a result of patient requests.
- Other methodological points as outlined above.

52. In summary, because of because of the existence of numerous confounders as alternative explanatory determinants of the effect of DTCA, significant limitations in the representativeness of the study participants, and in the extent and quality of the data they provide, I conclude that the Mintzes Study provides at best only weak evidence on the possible effects of DTCA. The particular conclusions drawn by the authors themselves cannot be reliably supported.

53. I make this affidavit in support of CanWest's application and for no other or improper purpose.

SWORN before me in the Town of Oakville, in the Province of Ontario this 29 day of May, 2007.

Stephen Walter

A COMMISSIONER

659996_1.DOC