Chapter 4. Limitations

We address the limitations of each set of analyses separately: meta-analysis of the weight loss and descriptive synthesis of athletic performance randomized controlled trials; analysis of the adverse events from the randomized controlled trials; and analysis of the case reports of adverse events.

The systematic reviews of the weight loss and athletic performance randomized controlled trials have the following potential limitations:

- Our search procedures for randomized controlled trials were extensive and included canvassing experts regarding studies we may have missed. In addition, we observed little to no evidence of publication bias via visual inspection or formal testing for the weight loss studies. However, we acknowledge that publication bias may still exist despite our best efforts to conduct a comprehensive search and the lack of statistical evidence of the existence of bias. Publication bias may occur for a variety of reasons, including investigators' loss of interest in the study if "negative" results are found or if results are obtained that are contrary to the interest of the sponsor or investigator.

- An important limitation common to many systematic reviews, whether or not a formal meta-analysis is conducted, is the quality of the original studies. Many of the weight loss studies suffered from an attrition rate higher than is normally allowed by FDA when assessing studies of pharmaceutical products seeking approval. However, recent attempts to define elements of study design and execution that are related to bias have shown that in many cases, such efforts are not reproducible and do not distinguish studies based on their results. Therefore, the current state of the science is to document such methodological weaknesses and perform sensitivity analyses when possible, but not to reject studies or use quality criteria to adjust the pooled outcome. We performed a sensitivity analysis on the subset of studies that had the best quality, according to the only validated scale available. The results of the sensitivity analysis did not alter the majority of our findings.

- In our meta-analysis of the weight loss studies, we did not observe significant evidence of heterogeneity. However, the chi-squared test of heterogeneity is underpowered. We did use a random effects approach to attempt to incorporate any heterogeneity and conducted sensitivity analyses to assess the robustness of our conclusions.

- We were limited by the small number of trials that provided direct comparisons between treatments of interest in the weight loss meta-analysis. Our meta-regression in this setting was an attempt to compare treatments across trials, but we acknowledge this approach does not allow for controlling for confounders within study. Direct comparisons are needed to draw more definite conclusions. In other words, while our observed results suggest that the amount of weight loss is approximately the same for ephedrine with caffeine, herbal ephedra with herbs containing caffeine, and herbal ephedra alone, the available data do not prove equivalence.

- The weight loss studies as a group had limited treatment duration; thus, we cannot draw conclusions about the association between ephedra or ephedrine and weight loss over longer and more clinically relevant intervals than about four months. Current knowledge of weight loss is that it generally ceases after six months, irrespective of treatment, and any weight loss is generally regained. Current recommendations for appropriate clinical trials in this area include a much longer treatment duration (at least one year) and an evaluation of what happens after the agent is withdrawn.

- The heterogeneity among the athletic performance studies prevented us from conducting a formal meta-analysis, so we were restricted to a descriptive synthesis.

- The results of the clinical trials are directly applicable only to the persons studied in those trials. In most cases, enrollment was highly selective to avoid certain comorbidities. Whether efficacy would be equivalent in a more representative population is unknown.

- The results of the ephedra studies regarding efficacy cannot be generalized to all ephedra-containing dietary supplements, because these may vary in their constituents from the concoctions studied and reported on here.

The analysis of the adverse events from the randomized controlled trials have the following major potential limitations:

- In this analysis, we focused only on studies that addressed weight loss or athletic performance. Although we observed no serious adverse events in these trials, we might have identified adverse events in trials that tested the efficacy of ephedra for other conditions, had we included those conditions in our search. However, we did include all controlled trials of ephedra or ephedrine for weight loss or athletic performance; therefore, our estimates are relevant to the populations taking those supplements for these reasons, which certainly constitute the majority of users of ephedrine and ephedra products in the United States.

- As with efficacy, the results of the clinical trials with respect to safety are directly applicable only to the persons studied in those trials. In most cases, enrollment was highly selective to avoid certain comorbidities. Whether safety is equivalent in a more representative population is unknown.

- As with efficacy, the results for the ephedra studies with respect to safety cannot be generalized to all ephedra-containing dietary supplements, because these may vary in their constituents from those concoctions studied and reported on here.

The analysis of the case reports of adverse events had the following major potential limitations:

- We did not have access to all adverse event files.

- Many authorities consider MedWatch case reports to underestimate the number of events, because patients need to suspect an association in order to report an event.

- This report did not review in detail other lines of evidence, such as animal studies, basic neuroscience studies, and adverse event data concerning other sympathomimetic amines that some authorities consider important when trying to assess causation.

- Many of the adverse event reports did not contain all the data that we needed to make assessments. Therefore, how the cases we classified as "insufficient evidence" might have influenced our findings had they contained appropriate documentation is unknown.

- An important limitation is that we do not have an estimate of the number of people using ephedra or ephedrine; that is, we do not have a denominator with which to calculate an event rate. An additional complication, we believe, is that the use of ephedra and ephedrine is increasing over time, as is the probability that someone will report an adverse event due to publicity.

- The most important limitation is that the study design (that is, an assessment of case reports) is insufficient for us to reach conclusions regarding causality.

The major potential limitations of the analysis of the Metabolife files can be classified into two categories: limitations of the source material and limitations of our methods.

The source material for this review differed in several important ways from source material used in other EPC projects:

- Much of what we reviewed was handwritten. Therefore, when the handwriting was poor we may not have correctly interpreted what the writer meant to say.

- The information was not recorded in an organized fashion, leaving it up to us to interpret its meaning. A good example of this was MIPER 23695 that we (but not Metabolife) classified as a "death." This file consisted of handwritten notes that stated, "migraine HA, wants refund, sister's husb died." Does this mean the customer is the sister's husband, who had a migraine headache and then died? Or did the customer have a migraine headache, perhaps in part because her sister's husband died? Without additional information it is impossible to tell.

- Each file did not attempt to collect the same information, so a recording bias probably exists.

- As already noted, we are not confident we could identify all files associated with a single case, so some double-counting may have occurred.

The methods we used to review the files also had important limitations:

- We relied on single-person review to screen cases. In the eight weeks we were given to review the files, we could not do dual review (which is standard in all our other EPC work) of over 18,000 cases. Therefore, more coding errors may have occurred than in situations where we use dual review. Mitigating this limitation is that we did do formal inter-rater reliability testing and demonstrated excellent reliability among reviewers. Also, the principal investigator reviewed all cases that were identified as serious. Furthermore, we identified nearly all the serious events identified by Metabolife, plus many more that Metabolife did not identify. So, while we acknowledge that there may still be errors in the data, we do not think they are so numerous or egregious as to threaten our conclusions.

- The Metabolife analysis did not undergo as extensive a review process as did the other sections of this report. The Metabolife analysis was reviewed by three experts and two federal agencies, in contrast to the much more extensive review process for the other sections of the report. Furthermore, because of the timeline necessary to produce the final report, the time available to the reviewers of the Metabolife analysis was shorter than we normally afford. How additional peer review may have affected our conclusions from the Metabolife analysis is unknown.