The final publication is available at link.spreinger.com via https://
doi.org/10.1007/s40279-017-0765-4

Appendix

July 6, 2017

Below, we provide supplementary materials for the article "Doping in Two Elite Athletics Competitions Assessed by Randomized-Response Surveys". These materials include the following:

Contents

List of Figures

List of Tables

1 Description of the Two Sports Events

The data collection took place in two separate and unrelated events, namely the 13th International Association of Athletics Federations (IAAF) World Championships in Athletics (WCA) and the 12th Quadrennial Pan-Arab Games (PAG). The WCA took place in Daegu, South Korea from August 27th, 2011 to September 4th, 2011. At this event, 1841 accredited athletes from 205 countries competed in 26 different track and field events.[1] The PAG took place in Doha, Qatar, from the December 6th, 2011 to December 23rd, 2011, and hosted 3346 accredited athletes from 21 countries, who competed in 31 events.[2] Tables 1 and 2 at the end of this document list the number of athletes from each country who participated in WCA and in PAG, respectively.

We obtained permission to distribute the survey at WCA from the International Association of Athletics Federations and permission to distribute the survey at PAG from The Union of Arab National Olympic Committees. General approval to perform anonymous surveys to assess the prevalence of substance use and other sensitive behaviors in public populations, using UQM (as defined in main next), was obtained through The Kingston University Science Faculty Research Ethics Committee (Andrea Petróczi, Principal Investigator).

2 Survey Instrument

2.1 Details of the Survey Instrument Administration

In this section, we describe and depict in detail the survey instrument that was used for the UQM at both WCA and PAG. The UQM question about doping was virtually identical at both of the two sporting events, but the control UQM question about use of supplements (described in the main text of this paper) was employed only at PAG. Therefore, for purposes of illustration, we first present the successive screens encountered by the athletes at PAG as they progressed through the instrument. Then, at the end of the section, we describe the differences between the instrument administered at WCA and that administered at PAG.

Figures 1–9 at the end of this document depict all nine successive images of the screens of the tablet computer seen by athletes when they performed the survey at PAG. The first screen asked the athlete to select his or her sport. This screen is shown in Figure 1, which depicts the upper part

[1] (i) International Association of Athletics Federations: IAAF World Championships, Daegu, Korea, 2011, available at https://en.m.wikipedia.org/wiki/2011_World_Championships_in_Athletics. (ii) International Association of Athletics Federations, personal communication, September 9, 2011.

[2] https://en.m.wikipedia.org/wiki/2011_Pan_Arab_Games.

of the screen in the left panel and the lower part of the screen in the right panel. As soon as the athlete selected his or her sport and touched the corresponding button on this screen, the tablet displayed the second screen. Alternatively, if athletes wanted to refuse participation, they could scroll down to the bottom of the first screen (right panel of Figure 1) and press the red button at the bottom of the screen to decline the survey. The second screen asked the athlete to select his or her language (Figure 2). Athletes at PAG could choose English, French, or Arabic, and all screens thereafter would then be displayed in the chosen language. Once the athlete had touched the language button, Screen 3 (Figure 3) appeared, introducing the survey, explaining that it was less than 2 minutes in duration, and stressing that respondents should be honest, since no one other than themselves knows which question they answered. Pressing the yes-button displayed Screen 4 (Figure 4), whereas pressing the no-button returned the athlete back to Screen 1.

Screen 4 asked the athlete to imagine a person whose date of birth was known to him or her. Upon touching the yes-button (indicating that they had mentally selected a "chosen person"), athletes were forwarded to survey page 1 (Figure 5). Individuals pressing the no-button were returned to Screen 1. Screen 5 explained the rules for determining which of two subsequent questions the athlete should prepare to answer on the following screen, namely "Question A" if the chosen person's birth date fell within the first 10 days of the month, or "Question B" if the chosen person's birth date fell between the 11th and 31st day of the month. Screen 5 again reminded athletes to answer honestly when they proceeded to their assigned question, and then provided a "Press to continue" button. Touching this button started a timer for measuring the athlete's response time and immediately displayed Screen 6 (Figure 6). Screen 6 reiterated the text of Screen 5 in its upper part and showed the nonsensitive Question A (an innocuous birth date question) and the sensitive Question B (the question regarding doping or supplement use) in its lower part. Once again, the screen encouraged honesty by reminding the athletes that only they themselves could know which of the two questions they were answering. Pushing the yes-button or no-button at the bottom of the screen stopped the timer. The tablet computer then recorded and stored the answer and the associated response time. Screens 7 and 8 (Figures 7 and 8) repeated this procedure for the second sensitive question (supplement use). The temporal order of the two sensitive questions (Screens 5 and 6 vs. Screens 7 and 8) was counterbalanced across athletes. In addition, the spatial positions (left and right) of the yes-button and the no-button were also counterbalanced across athletes. Statistical analyses revealed that these two spatial button arrangements did not exert an influence on the prevalence estimates for doping ($\hat{\lambda} = 53.2$ vs. 56.3%, $z = 0.98$, $p = 0.32$, by large-sample z-test, two-tailed, for left vs. right position of the yes-button, respectively).

The UQM questionnaire in the WCA survey was virtually identical to the one at PAG. The three major differences were that (a) the question on supplements (Screens 5 and 6) was replaced by another experimental survey technique that we wanted to pilot and which is described in Section 2.2 immediately below; (b) athletes could chose among 21 languages instead of three (i.e., English, French, Spanish, Russian, Mandarin, Arabic, Czech, German, Estonian, Greek, Croatian, Swahili, Norwegian, Polish, Portuguese, Romanian, Slovenian, Slovakian, Turkish, Japanese and Korean), and (c) Screen 1 was omitted.

2.2 Pilot Testing of the Experimental Single Sample Count Technique

In parallel with our survey using the UQM method reported here, we took advantage of both athletic events to conduct pilot tests of a recently proposed new fuzzy response method called *Single Sample Count* (SSC; Petróczi et al., 2011). The SSC technique is derived from the Unmatched Count Technique (Dalton, Wimbush, & Daily, 1994), in which one group of participants receives a sensitive target question (e.g., have you used illicit drugs?) embedded among k innocuous questions (e.g., do you like dogs?), while a control group of participants from the same population receives only the k innocuous questions without the sensitive question. Participants are asked only to report the number of questions that produced a "yes" answer, without disclosing which specific questions were positive. Thus, as is the case with other RRT methods described above, the investigators cannot deduce the status of any individual participant on the sensitive question, but they can calculate the estimated population prevalence of positive answers on the sensitive question by assessing the surplus number of "yes" answers reported by members of the sensitive-question group as compared to the number of "yes" answers reported by the control group.

The SSC uses a similar strategy, but improves over the Unmatched Count Technique in that it presents the sensitive question together with k innocuous questions that each have a known population prevalence of "yes" answers (e.g., were you born during the first six calendar months of the year?). Thus, the expected number of "yes" answers to the k innocuous questions can be calculated, and any surplus of "yes" answers may be assumed to reflect positive answers to the sensitive question. With the SSC method, all participants receive the sensitive question and no control sample is required, making data collection more economical. During the present study, we performed pilot testing with an experimental version of this new estimation model to assess its possible advantages over traditional random response models, including its potential ability to model the magnitude of noncompliance, and further to estimate the proportion of noncompliance attributable to strategic self-protection, as opposed to noncompliance arising from lack of motivation or failure to understand the instructions.

We administered this experimental version to all athletes surveyed at WCA in Korea, such that where athletes received both the established method described in this paper (UQM) and the experimental method (SSC) in counterbalanced order; and at PAG where athletes were randomly assigned to one of the two models (Table 6 displays the frequency counts obtained with the SSC method). A more comprehensive discussion of the SSC methods for adjusting for noncompliance, together with the SSC-derived and noncompliance-corrected doping estimations from these field tests, will be presented in a subsequent publication.

3 Basic Results

As described in the main text, we approached 1290 (70.1%) of the 1841 registered athletes at WCA, and of these, 1203 (93.3%) agreed to participate. The data of one athlete had to be discarded because of a machine error and thus the data of the remaining 1202 athletes were analyzed. At WCA, we employed six data collectors (three men and three women), chosen because they were all themselves athletic enthusiasts, and who collectively spoke 10 languages (English, French, German, Spanish, Italian, Arabic, Russian, Korean, Mandarin, and Cantonese). The data collectors cooperated extensively to ensure that they were reaching athletes from all sports disciplines and from all countries participating in the games. The data collectors provided a brief verbal explanation of the project upon presenting the tablet computer to each athlete, noting that the survey would require only two minutes and was designed to guarantee the confidentiality of the athlete's responses. Athletes who performed the survey received a punch hole in their identification tags (which they were required to wear at all times), thus allowing data collectors to identify and target athletes who had not yet participated, while avoiding duplication of athletes who had already participated. Athletes who declined to perform the survey after being approached also received a punch hole so that they would not be approached again. Athletes who completed the survey received a token reward (a combined flashlight/whistle) in appreciation of their cooperation. The number of athletes approached was limited only by the availability of data collectors and tablet computers, and the subgroup of participating athletes appeared representative of the entire group of registered athletes. Of note, essentially 100% of registered athletes at WCA lived in the athletes' village constructed for the event, and ate regularly at the cafeteria in the village. The six data collectors approached athletes throughout the village, including especially the cafeteria, throughout the days during the course of the games.

Further assurance regarding the representativeness of the sample is provided by examination of the languages chosen by the athletes. All of the 21 languages offered on the WCA survey (see the list at the end of Section 2.1) were utilized, and the portions of athletes utilizing different languages appeared to correspond well with the sizes of the various athlete delegations from different countries. Of course, any analysis based on languages is only approximate, since athletes might not always choose the native language of their country, and conversely many of the languages offered on the survey were spoken by athletes from more than one country. However, looking at several representative languages that were relatively specific to a single country, we noted that 27 (2.2%) of the 1202 participating athletes chose to perform the survey in Mandarin Chinese, and WCA data showed that the Chinese delegation plus the Chinese Taipei delegation collectively comprised 60 (3.3%) of the total pool of 1841 registered athletes. Looking at several other languages that were relatively specific to a single country, 0.8% our respondents chose Greek, and 0.7% of registered athletes were from Greece; 2.7% chose Japanese, and 2.7% of all athletes were from Japan; 3.7% chose Korean, and 3.3% of all athletes were from Korea; 2.5% chose Polish, and 2.3% of all athletes were from Poland; 0.8% chose Turkish, and 1.1% of all athletes were from Turkey. Although it must be reemphasized that these comparisons provide only a crude approximation, due to the imperfect correspondence between languages and countries, they nevertheless suggest that the WCA sample was generally representative of the entire pool of athletes.

At PAG, we employed four data collectors (one of whom was LC, an author of this paper), who collectively spoke English, French, Spanish, and Arabic. Of the 3346 athletes accredited at the PAG, we approached 2050 (61.3%). The athletes at PAG received either the experimental SSC (described in Section 2.2 above) or the standard UQM presented in this study in counterbalanced order. Of 1020 athletes receiving the SSC, 954 (93.5%) agreed to participate; of 1030 athletes receiving the UQM, 965 (93.6%) agreed to participate, and thus the data of these 965 athletes were analyzed. As at WCA, the number of athletes approached was limited only by the availability of data collectors and tablet computers, and the subgroup studied appeared representative of the entire population of accredited athletes. Once again, virtually 100 percent of the athlete population was housed at the athletes' village in Doha, and essentially 100% of athletes ate at the cafeteria there, thus ensuring that the data collectors had access to the entire population. A similar method of punch holes in the athletes' identification tags was used to distinguish athletes who had completed the survey (or who had declined after being approached) from those who had not. Athletes completing the survey received a flashlight/whistle token similar to that provided at the WCA.

Unlike the WCA survey, the PAG survey utilized only three languages, as described in Section 2.1, so that it was not possible to use the language distribution to gauge the representativeness of the sample as described for WCA above. Moreover and also in contrast to the WCA survey, athletes in the PAG survey were additionally asked to indicate the particular sport in which they were competing (see tablet screen 1 in Figure 1 below). Thus we could compare the distribution of the sports cited by survey respondents with the distribution of sports for the entire population of athletes at the games (see Table 3). Again, this comparison suggested that our sample was representative, with survey respondents in every PAG category of sports, in proportions generally similar to the proportions for these categories in the overall population of athletes at PAG. Comparing these proportions and looking at the most extreme discrepancies in each direction, we found that athletics was the sport most overrepresented in our sample (14.8% of our respondents versus 9.2% of PAG athletes overall) and paralympics the most underrepresented (0.4% of our respondents versus 4.1% of athletes overall). However, the latter discrepancy may simply reflect the fact that paralympic contenders likely preferred to classify themselves by the sport in which they competed, rather than by the fact that they were disabled and hence defined as "paralympic" in the official PAG nomenclature.

The major findings from our primary analysis of the UQM results are summarized in Table 4. This table shows the observed percentage $\hat{\lambda}$ of "yes" answers for the past-year doping items at WCA and PAG and the past-year supplements item at PAG . (Note that $\hat{\lambda}$ represents the raw percentage of "yes" answers for each item, without knowledge of whether the athlete was answering "yes" to Question A or B on that item). For each of these behavioral items, the table also provides the estimated prevalence $\hat{\pi}_S$, its standard error SE, and its 95% confidence interval, calculated on the basis of Equations 1 and 2 presented in the main text.

For each athlete we also measured response time, defined as the interval from the onset of the display containing the Questions A and B to the time that the respondent pressed the "yes" or "no" button on the tablet in response to one of these questions, as explained above. Figure 10 displayes histograms for these response times and for every question. The mean (SD) response time was 20.2 (17.9) sec for the doping item at WCA, 20.2 (20.4) sec for doping at PAG, and 18.7 (18.1) sec for supplement use at PAG. We then considered the possibility that athletes with very fast response times might produce unreliable answers as a result of haste or carelessness (as discussed in the main text of this paper). Therefore, for each of the three items, we divided the athletes' response times into deciles, and then performed sensitivity analyzes in which we progressively deleted from the total population those deciles of athletes showing the fastest responses (e.g., the fastest 10%, 20%, 30%, 40%, and 50%). The three parts in Table 4 contain the results of these sensitivity analyses for each question. These results are also summarized in Figure 1 in the main text.

We next examined whether the presentation order of the item had influenced the prevalence estimates. In the PAG study, the doping item was presented either before or after the supplements item, and in the WCA study, the doping item was presented before or after the experimental SSC item (described in Section 2.2 above). Table 5 provides the results of this analysis. (Table 6 gives the frequency counts for the SSC method when it was presented first or second). For all of the three items, we found that the temporal position of the item had no significant effect on the prevalence estimates. Specifically, comparing the number of "yes" answers among athletes receiving a given item first versus those receiving it second, using a large-sample z-test to compare the proportions of two independent samples, we found no significant sequence effect for the doping item at WCA ($z = 1.18$, $p = 0.24$, two-tailed), the doping item at PAG ($z = 0.07$, $p = 0.94$), or the supplements item at PAG ($z = 1.16$, $p = 0.25$). These observations suggest that no carryover effect occurred when answering the second question after having processed the first one.

By contrast, consistent with the power law of practice (Newell & Rosenbloom, 1981), we found that the temporal position of an item significantly affected athletes' response time for that item. Specifically, comparing athletes receiving a given item first versus those receiving it second, again using a large-sample z-test, we found that response time when receiving the item second was significantly shorter for doping at WCA $z = 3.60$, $p < .001$, doping at PAG, $z = 7.93$, $p < .001$, and supplements at PAG, $z = 7.93$, $p < .001$. It was not surprising to find that the effect of practice was significantly greater for doping at PAG (mean RT difference 10.0 sec) than for doping at WCA (mean RT difference 3.7 sec; $z = 2.28$, $p = 0.012$, one-tailed). This was because at PAG, athletes who received the doping item second had already performed the supplements item in the identical UQM format, and thus were practiced with this format. By contrast at WCA, athletes who received the UQM doping item second had previously performed the experimental SSC technique, and hence had no prior practice on the UQM when they encountered it. Thus the modest 3.7 sec improvement in speed at WCA presumably reflected only a nonspecific effect of practice (e.g., becoming familiar with the tablet presentation in general).

4 Effects of Various Possible Forms of Noncompliance by Survey Respondents

We next discuss the effects of various possible forms of respondent noncompliance that might have biased the prevalence estimates generated by the UQM. The sections below address seven different possible forms of noncompliance in order to assess the robustness of the prevalence estimates generated by the UQM under each of these scenarios.

4.1 Effects of Random Responding

One might argue that a certain proportion c of athletes in haste did not process the instructions or other information properly, and thus responded randomly with "yes" or "no", regardless of doping status. Under this scenario, the proportion of observed "yes" responses is given by

$$\lambda = c \cdot 0.5 + (1 - c) \cdot [p \cdot \pi_S + (1 - p) \cdot 0.5]. \tag{1}$$

Solving the preceding equation for π_S, yields

$$\pi_S = \frac{\lambda - (1-p) \cdot 0.5 - c \cdot p \cdot 0.5}{p \cdot (1-c)}. \tag{2}$$

Table 7 shows how the prevalence estimates reported in the main text would change when c is assumed to be equal to 10, 20, or 30%. As can be seen, even if we were to assume that as many as 30% of the athletes who participated at WCA and at PAG had responded randomly, this assumption would still only slightly change the prevalence estimates from those obtained under the standard UQM assumption that $c = 0$.

4.2 Effects of Underreporting of Doping

Despite the assurances of confidentiality provided by the randomized UQM method, some dopers who received the sensitive Question B might nevertheless have lied as a self-protective measure, and hence responded with "no" instead with "yes". This type of possible noncompliance was recently discussed in conjunction with Warner's RRT model by Moshagen, Musch, and Erdfelder (2012), and the possibility of such noncompliance must also be considered with the UQM. Therefore, let us assume that the proportion of such dishonest dopers was equal to c. In this noncompliance scenario, the portion of "yes" responses is given by

$$\lambda = p \cdot \pi_S \cdot (1-c) + (1-p) \cdot 0.5 \tag{3}$$

and rearranging this equation yields

$$\pi_S = \frac{\lambda - (1-p) \cdot 0.5}{p \cdot (1-c)}. \tag{4}$$

Table 8 demonstrates how our prevalence estimates in the main text would change if the proportion c of dishonest dopers (e.g., dopers who received Question B and falsely answered "no") were equal to 10, 20, or 30%. As one might expect, our standard UQM calculation, with its assumption that $c = 0$, would have strongly underestimated the true prevalence of doping within our samples if the proportion of dishonest dopers were 10% or higher. Thus, the standard UQM estimate likely provides a conservative lower bound of the true prevalence of doping, which might be substantially greater if there were even modest numbers of dishonest dopers.

4.3 Effects of Switching from Question B to Question A

Another noncompliance scenario presumes that some athletes who were directed to the sensitive Question B might have conceivably switched back to the nonsensitive Question A because they did not wish to answer Question B. In other words, on the opening screen, an athlete might have chosen a "close person" who was born after the first 10 days of the month, thereby requiring that athlete to answer Question B on the following screen. However, upon being confronted with the sensitive content of Question B, that athlete might have cheated and retroactively chosen a new and different "close person" who was born during the first 10 days of the month, in order to be entitled to answer nonsensitive Question A instead. More specifically, assume that a proportion of g_D dopers who had received Question B switched back to A; likewise let the proportion of non-dopers who switched from B to A be equal to g_N. It can be shown that under this noncompliance scenario, the proportion of total "yes" responses is given by

$$\lambda = p \cdot \pi_S \cdot (1-g_D) + p \cdot \pi_S \cdot g_D \cdot 0.5 + p \cdot (1-\pi_S) \cdot g_N \cdot 0.5 + (1-p) \cdot 0.5. \tag{5}$$

After solving for π_S and simplifying, one obtains

$$\pi_S = \frac{\lambda - (1 - p + p \cdot g_N) \cdot 0.5}{p \cdot [1 - (g_D + g_N) \cdot 0.5]}. \tag{6}$$

Intuitively, it would seem that dopers might be motivated to switch from Question B back to Question A in order to avoid having to answer the sensitive question positively, whereas non-dopers would have little motivation to switch backwards, since their answer to Question B would already be "no", and thus they would have nothing to gain by switching back to Question A. Therefore the most likely scenario would assume that most or all of the athletes who switched from B to A would be dopers. Nevertheless, we consider below three different possibilities of g_D and g_N, extending even to seemingly unlikely scenarios in which substantial portions of non-dopers are assumed to have switched despite the absence of apparent motivation. These possibilities are: (a) only dopers switch, i.e., $g_D > 0$ and $g_N = 0$, (b) more dopers than non-dopers switch, i.e., $g_D = 2 \cdot g_N$, and (c) an equal number of dopers and non-dopers switch, i.e., $g_D = g_N$. Parts A, B, and C in Table 9 show the estimated true prevalence within our sample if we assume that these various noncompliance scenarios had occurred, for various proportions of switching (0-30%), respectively. As can be seen from the numerical examples, the effect of switching would not meaningfully distort of the estimates obtained under the assumptions of the standard UQM. If dopers had switched more often than non-dopers (i.e., scenarios a and b), the true prevalence of doping would be meaningfully higher than our estimate in the main text, based on the $c = 0$ assumption of the standard UQM. In the seemingly less plausible scenario that non-dopers were just as likely to switch as dopers (scenario c), the true prevalence of doping within our sample would still differ only slightly from our estimates under the standard UQM assumption.

4.4 Effects of Switching from Question A to Question B

In contrast to the previous scenario, some athletes who were directed to Question A might have conceivably switched to the sensitive Question B. For example, it is plausible that some non-dopers, upon receiving Question A and finding that their answer to this question was a "yes", might have been motivated to switch from Question A to B so that they would be enabled to answer "no" and thus stress their non-doping status. However, it is also conceivable that the converse situation might have occurred, namely that some dopers, upon receiving Question A and finding that their answer to this question was a "no", might have been motivated to switch from Question A to B so that they would be enabled to answer "yes" and therefore be able to confess doping in such a confidential situation.

Thus let g_D be the proportion of all dopers who had received Question A and switched to B; likewise let g_N be the proportion of all non-dopers who switched from A to B. Under this noncompliance scenario, one obtains

$$\lambda = p \cdot \pi_S + (1-p) \cdot \pi_S \cdot [g_D + (1-g_D) \cdot 0.5] + (1-p) \cdot (1-\pi_S) \cdot (1-g_N) \cdot 0.5 \tag{7}$$

and thus

$$\pi_S = \frac{\lambda - (1-p) \cdot (1-g_N) \cdot 0.5}{p + (1-p) \cdot (g_D + g_N) \cdot 0.5}. \tag{8}$$

Paralleling our approach in the preceding section, Parts A, B, and C in Table 10 illustrate three possibilities of g_D and g_N, ranging from the scenario in which only non-dopers switched from A to B, and extending to the scenario in which substantial portions of dopers are assumed to have switched from A to B: (a) only non-dopers switch, i.e., $g_N > 0$ and $g_D = 0$, (b) more non-dopers than dopers switch, i.e., $g_D = 2 \cdot g_N$, and (c) an equal number dopers and non-dopers switch, i.e., $g_D = g_N$. As can be seen in Table 10, none of these scenarios would result in marked changes in the prevalence estimates.

4.5 Effects of Responding to Question A or B Regardless of the Instructions

Another noncompliance scenario assumes the possibility that some athletes, regardless of their doping status, might have ignored the instructions and simply jumped to Question A (e.g., perhaps because they were in haste and simply answered the question on the top). If c is the proportion of athletes that exhibited this behavior, then the predicted total proportion of “yes” responses is

$$\lambda = c \cdot 0.5 + (1-c) \cdot [p \cdot \pi_S + (1-p) \cdot 0.5]. \tag{9}$$

Note that this expression is identical to Equation 2 derived for the Random-Responding Model. Therefore, the scenario of ignoring the instructions and jumping to Question A would have produced the same results shown in Table 7 for various proportions of athletes assumed to have engaged in this behavior.

We next consider the alternate possibility that some athletes might have ignored the instructions and jumped directly to Question B rather than to Question A. Again let c be the number of athletes who ignored the instructions and immediately jumped to Question B. In this case the predicted total proportion of “yes” responses is

$$\lambda = c \cdot \pi_S + (1-c) \cdot [p \cdot \pi_S + (1-p) \cdot 0.5]. \tag{10}$$

Rearranging this expression leads to

$$\pi_S = \frac{\lambda - (1-c) \cdot (1-p) \cdot 0.5}{c + (1-c) \cdot p}. \tag{11}$$

Table 11 contains the prevalence estimates under the assumption that 0, 10, 20 or 30% of the athletes ignored the instructions and simply jumped to Question B. It can be seen that such behavior, if it occurred, would only slightly alter the estimates obtained under the standard UQM.

4.6 Effects of Automatic Responding Regardless of the Question Received

In this section, we examine how automatic or hasty responding might bias UQM prevalence estimates. Cognitive research has documented that responses which are triggered by automatic processes are especially fast in contrast to responses that are triggered by deliberative processes (e.g., Kahneman, 2011; Moors & De Houwer, 2006). It is conceivable that the responses by some athletes might have been triggered by such fast automatic processes.

Therefore, as already described in the main text and also in Section 2 above, we performed an analysis in which we incrementally deleted subgroups of fast responders, in order to exclude athletes who may have responded automatically (cf. Table 4). As noted in the main text, the prevalence estimates generated by the UQM declined with the exclusion of the fastest responders, but reached a plateau and showed no further declines after approximately 30% of the fastest responders were excluded. It appears likely that the exclusion of the fastest 30% of responders would almost certainly eliminate individuals who responded hastily and automatically, thus generating a group no longer contaminated by automatic responding.

Nevertheless, it is important to assess the potential effects of automatic responding. First, we consider the effect of automatic “yes” responding and then the effect of automatic “no” responding.

4.6.1 Automatic “yes” Responding

Beginning with the scenario of automatic “yes” responding, we assume that dopers and non-dopers automatically respond “yes” with probability a_D and a_N, respectively. With the complementary probabilities $1-a_D$ and $1-a_N$, dopers and non-dopers respond deliberately and non-automatically in proper accordance with the instructions. Under this assumption one obtains

$$\begin{aligned}\lambda &= \pi_S \cdot \{a_D + (1-a_D) \cdot [p + (1-p) \cdot 0.5]\} \\ &\quad + (1-\pi_S) \cdot \{a_N + (1-a_N) \cdot (1-p) \cdot 0.5\}\end{aligned} \tag{12}$$

and thus after rearranging and simplifying

$$\pi_S = \frac{2 \cdot \lambda - a_N \cdot (1+p) - (1-p)}{a_D \cdot (1-p) - a_N \cdot (1+p) + 2 \cdot p}. \tag{13}$$

In particular, if one assumes that the extent of automaticity is about the same for dopers and non-dopers (viz. $a_D = a_N = a$), the preceding equation further simplifies to

$$\pi_S = \frac{\lambda - [a \cdot (1+p) + (1-p)] \cdot 0.5}{p \cdot (1-a)}. \tag{14}$$

Table 12 gives the prevalence estimates generated under the assumptions that 0, 10, 20 or 30% of the athletes automatically responded with "yes" irrespective of whether they were dopers or not and regardless of the question received. Under this scenario, UQM would strongly overestimate the true prevalence rate.

4.6.2 Automatic "no" Responding

Automatic "no" responding is also a plausible alternative, especially if athletes are particularly eager to avoid creating the impression that their sport or discipline is prone to doping. In this case, one would expect that both dopers and non-dopers would respond with "no" regardless of whether they received the sensitive or neutral question.

Thus in this scenario, we assume that dopers and non-dopers automatically respond "no" with probability a_D and a_N, respectively, with complementary probabilities $1 - a_D$ and $1 - a_N$, that dopers and non-dopers would respond deliberately and non-automatically in proper compliance with the instructions. Under this assumption one computes

$$\begin{aligned} \lambda &= \pi_S \cdot (1-a_D) \cdot [p + (1-p) \cdot 0.5] \\ &\quad + (1-\pi_S) \cdot (1-a_N) \cdot (1-p) \cdot 0.5 \end{aligned} \tag{15}$$

and obtains after rearranging and simplification

$$\pi_S = \frac{2 \cdot \lambda - (1-p) \cdot (1-a_N)}{a_N \cdot (1-p) - a_D \cdot (1+p) + 2 \cdot p}. \tag{16}$$

In particular, under $a_D = a_N = a$, the preceding equation reduces to

$$\pi_S = \frac{\lambda/(1-a) - (1-p) \cdot 0.5}{p}. \tag{17}$$

Table 12 also gives the prevalence estimates generated under the assumptions that 0, 10, 20 or 30% of the athletes automatically responded with "no". Under this scenario, however, UQM would strongly underestimate the true prevalence rate.

4.7 Effects of Cheating with "no" Regardless of the Question Received

Some athletes may have believed that their privacy was not sufficiently protected and thus intentionally responded with "no" regardless of the questions. Such "no" responding would seem most likely to occur among dopers, but one could envisage that even some non-dopers might have used this strategy, simply to avoid any impression that they were engaged in doping — a possibility proposed by Clark and Desharnais (1998) and acknowledged by several RRT researchers (e.g., Böckenholt & van der Heijden, 2007).[3] Therefore, we assume here that the athletes who ignored the instructions and responded with "no" might have included both non-dopers and dopers.

In order to derive the predictions for this scenario within the framework of UQM, we divide the whole sample into cheaters ("no-sayers") and non-cheaters. More specifically, let $P(\text{Cheater}) = c$ denote the probability that a given athlete is a cheater and the complementary probability,

[3] This version is similar to the preceding model. However, the basic assumption in this version is comparable to the cheater detection model by Clark and Desharnais (1998) and thus provides an alternative perspective of "no" responding.

$P(\text{Non-Cheater}) = 1 - c$, denote that he or she is honest. Furthermore, let $P(\text{Doper}|\text{Cheater}) = \pi_{S|C}$ denote the conditional probability that a given cheater is a doper.

Likewise let $P(\text{Doper}|\text{Non-Cheater}) = \pi_{S|N}$ represent the conditional probability that a non-cheater is a doper. With these definitions, the proportion π_S of dopers in the whole sample can be derived on the basis of probability theory as follows:

$$\begin{aligned}
\pi_S &= P(\text{Doper}) && (18)\\
&= P(\text{Doper} \cap \text{ Cheater}) + P(\text{Doper} \cap \text{ Non-Cheater}) && (19)\\
&= P(\text{Doper}|\text{Cheater}) \cdot P(\text{Cheater}) \\
&\quad + P(\text{Doper}|\text{Non-Cheater}) \cdot P(\text{Non-Cheater}) && (20)\\
&= \pi_{S|C} \cdot c + \pi_{S|N} \cdot (1 - c). && (21)
\end{aligned}$$

In other words, under this scenario the desired prevalence π_S represents the weighted average of the conditional probabilities $\pi_{S|C}$ and $\pi_{S|N}$.

Furthermore, according to the above assumptions, the proportion of "yes" responses is given by

$$\lambda = (1 - c) \cdot p \cdot \pi_{S|N} + (1 - c) \cdot (1 - p) \cdot 0.5. \quad (22)$$

Solving Equation 21 for $(1 - c)$ and inserting the resulting expression into the second term of Equation 22 results in

$$\begin{aligned}
\lambda &= \frac{\pi_S - \pi_{S|C} \cdot c}{\pi_{S|N}} \cdot p \cdot \pi_{S|N} + (1 - c) \cdot (1 - p) \cdot 0.5 && (23)\\
&= \pi_S \cdot p - \pi_{S|C} \cdot c \cdot p + (1 - c) \cdot (1 - p) \cdot 0.5. && (24)
\end{aligned}$$

If one solves the preceding expression for π_S and rearranges its terms properly, the final result is:

$$\pi_S = \frac{\lambda - (1 - p) \cdot 0.5}{p} + c \cdot \frac{(\pi_{S|C} - 0.5) \cdot p + 0.5}{p}. \quad (25)$$

Note that the first term on the right side of the last expression corresponds to the standard UQM prevalence estimate that assumes no cheating (compare Equation 1 in the main text). Also note that the second term on the right side can never be less than zero. Therefore, the standard estimate of the UQM would always underestimate the true prevalence within our sample when the automatic no-saying scenario applies. It can also be seen that the size of this bias increases with c and with $\pi_{S|C}$. For example for $\pi_{S|C} = 0.5$, $c = 30\%$, and $p = 2/3$, a bias of 0.225 is obtained. In other words, the prevalence estimate of the standard UQM would underestimate the true prevalence by 22.5 percentage points in this case.

4.8 Effects of Finger Errors

In the previous sections, we have considered how various scenarios of noncompliance could bias the prevalence estimates of the standard UQM. However, we should also consider the possibility that an athlete might have intended to follow the instructions but simply touched the wrong response button by accident. Although such finger errors (also called lapses) are most likely rare (i.e., less than 10%; see for example, Treutwein & Strasburger, 1999), one cannot exclude them either (Bausenhart, Dyjas, Vorberg, & Ulrich, 2012). Accordingly, let us assume that a finger error occurs with probability c regardless of the intended response and regardless of the athlete's doping status. Under this assumption one computes

$$\lambda = p \cdot \pi_S \cdot (1 - c) + (1 - \pi_S) \cdot c + (1 - p) \cdot [0.5 \cdot c + 0.5 \cdot (1 - c)] \quad (26)$$

and

$$\pi_S = \frac{\lambda - (1 - p) \cdot 0.5 - p \cdot c}{p \cdot (1 - 2 \cdot c)}. \quad (27)$$

Assuming $c = 10\%$, the last expression yields $\hat{\pi}_S = 41.1\%$, 58.9%, and 75.3% for doping (WCA), doping (PAG), and supplements (PAG), respectively. Therefore, it can be seen that finger errors would only slightly distort the prevalence estimates from the standard UQM.

4.9 Effects of Misinterpreting the Doping Question

It is possible that some athletes might have misinterpreted the doping question ("Have you knowingly violated anti-doping regulations by using a prohibited substance or method in the past 12 months?") which could lead to misclassification by producing either false-positive responses (i.e., erroneously believing that they were doping, and hence answering "yes" to the doping question, when in fact they had not violated doping regulations within the past year) or false-negative responses (erroneously believing that they were not doping when in fact they had violated regulations).

Several scenarios might plausibly yield false positive responses. First, athletes might believe that a substance was banned when in fact it was not. For example, meldonium was not banned in 2011 at the time of the study, but was eventually banned about five years later in 2016. Conceivably, some athletes taking meldonium as early as 2011 might have perceived it as a potentially prohibited or unethical substance, and hence answered "yes" on the doping question. Second, some substances are banned in competition but are permitted out of competition (e.g., cannabis used for recreation or central nervous system stimulants used in training). Some athletes who had used such substances, but only when out of competition, might have failed to recognize that such use was permissible, and hence erroneously claimed that they were doping. Third, athletes using drugs under a therapeutic use exemption (TUE) might have erroneously responded that they were doping when in fact their drug use was permitted under the TUE. However, this last scenario is likely rare: in a survey of 56 elite athletes currently granted a TUE, Overbye and Wagner (2013) assessed responses to a number of statements, including the statement, "I feel like it is cheating quite a bit." In answer to this statement, none of the athletes responded "corresponds completely," and only 1 (2%) responded "corresponds fairly well." Thus, elite athletes granted a TUE would seem very unlikely to erroneously perceive themselves as violating doping regulations.

Several alternative scenarios could yield false-negative responses. First, athletes might have used supplements containing banned substances but failed to recognize that these supplements represented doping violations, and hence erroneously answered "no" on the doping question. This possibility is supported by data indicating that many athletes may not be aware of all of the active ingredients in supplements that they are ingesting (Dascombe, Karunaratna, Cartoon, Fergie, & Goodman, 2010; Maughan, Depiesse, & Geyer, 2007). Second, athletes might have taken medications, with or without a doctor's prescription, while failing to recognize that these medications represented banned drugs. Third, athletes might have erroneously believed that drugs that they were taking were exempted because these drugs had been prescribed for therapeutic purposes. But if such athletes had failed to obtain a formal therapeutic use exemption, they would be erroneously denying doping when in fact they were violating doping regulations.

The authors believe, based on their collective knowledge of international athletics as well as the existing body of scientific knowledge, that these various false-positive and false-negative scenarios are rare. World-class athletes typically have extensive experience in numerous prior competitions, and are well educated both by anti-doping authorities and by sports federations with regard to doping regulations. Thus, our explicitly worded question ("Have you knowingly violated anti-doping regulations by using a prohibited substance or method in the past 12 months?") would be unambiguous to these sophisticated competitors.

Moreover, even if we assume for purposes of discussion that the athletes did exhibit substantial levels of naïveté, and frequently misinterpreted the doping question, the results of the present study would still not be seriously affected unless there was reason to believe the number of false-positive responses greatly outweighed the number of false-negatives or vice versa. However, it seems much more likely in this study that any potential misclassification was nondifferential (i.e., that any athletes' misunderstanding of the question occurred to about the same degree in both directions) and thus any net bias would be small.

Finally, even if we were to assume, hypothetically, that misinterpretation of the doping question somehow caused a substantial differential misclassification large enough to cause error in our estimates of the true prevalence of doping, our findings would still nevertheless represent an estimate of the number of athletes who believed that they were knowingly violating doping regulations. Thus, our findings would be of concern even in the unlikely scenario that the number of athletes who consciously believed that they were violating regulations differed substantially from the number who were technically violating regulations.

4.10 Summary and Evaluation of the Effects of Various Possible Forms of Noncompliance

In conclusion, we summarize the effects of all of the above possible forms of noncompliance, while noting that some of these scenarios are more plausible than others, as we have previously discussed.

Table 13 logically classifies all possible forms of noncompliance that we can envisage; the reader can confirm that each of these has been discussed in the main text and/or the supplemental online material above. The only scenarios arguably not embraced by Table 13 are those that do not fall under the heading "compliance," such as totally random responding (see Section 4.1), finger slips (Section 4.8), and positive or negative misinterpretation of the doping question (Section 4.9). However, as we have shown in these three sections, these three possible sources of error would seem unlikely to have substantially distorted our estimates.

Upon examining Table 13, it will be seen that most forms of noncompliance would have caused our estimate to underestimate the true prevalence of doping within our sample. Indeed, there are only four scenarios that could have caused us to overestimate the prevalence. Three of these four seem highly implausible: (a) non-dopers receiving Question B and falsely self-incriminating by answering "yes"; (b) non-dopers receiving Question B and switching to Question A despite any obvious motivation to do so; and (c) dopers receiving Question A and instructed to answer "no" to this question, but nevertheless answering "yes" (e.g., by switching to Question B). Moreover, the latter two of these possibilities would not markedly affect our estimates even if they did occur (Sections 4.3 and 4.4, respectively).

This leaves automatic "yes" responding as the only remaining possibility. Although this scenario would theoretically cause substantial overestimates of the prevalence of doping, this is not a serious problem because our exercise of deleting up to 30% of fast responders (see Section 3) almost certainly excluded virtually all of the automatic "yes" responders. Thus our "pure" samples of the remaining 70% of responders would presumably not be vulnerable to error from automatic "yes" responding. Our exercise would also presumably eliminate virtually all of the automatic "no" responders as well, and elimination of these individuals would of course work in the opposite direction and cause our prevalence estimate to be overly conservative. Therefore, our estimates generated on the basis of excluding 30% of the fastest responders in each case would likely represent a lower bound for the estimated prevalence of doping at each event.

To enable a summary evaluation of the possible effects of noncompliance, we quantified the potential biases for the various forms of noncompliance. Thus we computed the magnitude of bias for the various prevalence estimates, as shown in Tables 7–12, as the difference

$$D = \hat{\pi}_S(UQM) - \hat{\pi}_S(NON) \tag{28}$$

where $\hat{\pi}_S(UQM)$ denotes the prevalence estimate from the standard UQM as presented in the main text here, and $\hat{\pi}_S(NON)$ the estimate under the assumption that a given form of noncompliance occurred. A positive value of D indicates that our standard UQM estimate overshot the true prevalence within our sample, whereas a negative value of D indicates that our estimate undershot the true prevalence. In a final step, we classified the degree of bias D (i.e. the magnitude of the difference between our standard estimate and the estimate based on an assumption of noncompliance), D, as (a) $D \geq 3\%$, (b) $-3\% < D < 3\%$, or (c) $D \leq -3\%$. These three categories of bias were based on the assumption of the highest level of noncompliance considered in each of our tables above, namely that 30% of athletes engaged in the noncompliant behavior in question (usually the bottom row of the table).

The three right-hand columns in Table 14 provide these categories of bias separately for each posited form of noncompliance and for each behavior assessed (doping at WCA, doping at PAG, and supplements at PAG). As can be seen, none of the various assumptions would have caused us to overestimate the true prevalence within our sample by more than 3%; several assumptions could have affected our estimates by less than 3% in either direction, and many assumptions could have caused us to underestimate the prevalence by more than 3% (and often much more than 3%, as illustrated in the preceding tables).

In summary, it appears unlikely that we seriously overestimated, and more likely that we underestimated, and perhaps even markedly underestimated, the true prevalence of past-year doping among elite athletes at both events.

References

Bausenhart, K. M., Dyjas, O., Vorberg, D., & Ulrich, R. (2012). Estimating discrimination performance in two-alternative forced-choice tasks: Routines for MATLAB and R. *Behavior Research Methods*, *44*, 1157-1174.

Böckenholt, U., & van der Heijden, P. G. M. (2007). Item randomized-response models for measuring noncompliance: Risk-return perceptions, social influences, and self-protective responses. *Psychometrika*, *72*, 245-262.

Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods*, *3*, 160-168.

Dalton, D., Wimbush, J., & Daily, C. (1994). Using the unmatched count technique (UCT) to estimate base-rates for sensitive behavior. *Personnel Psychology*, *47*, 817–828.

Dascombe, B. J., Karunaratna, M., Cartoon, J., Fergie, B., & Goodman, C. (2010). Nutritional supplementation habits and perceptions of elite athletes within a state-based sporting institute. *Journal of Science and Medicine in Sport*, *13*, 274–280.

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Maughan, R. J., Depiesse, F., & Geyer, H. (2007). The use of dietary supplements by athletes. *Journal of Sports Sciences*, *25*, S103–S113.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297–326.

Moshagen, M., Musch, J., & Erdfelder, E. (2012). A stochastic lie detector. *Behavior Research Methods*, *44*, 222-231.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Overbye, M., & Wagner, U. (2013). Between medical treatment and performance enhancement: An investigation of how elite athletes experience Therapeutic Use Exemptions. *International journal of drug policy*, *24*, 579–588.

Petróczi, A., Nespusz, T., Cross, P., Taft, H., Shah, S., Deshmukh, N., Schaffer, J., Shane, M., Adesanwo, C., Barker, J., & Naughton, D. P. (2011). New non-randomised model to assess the prevalence of discimination behaviour: A pilot study on mephedrone. *Substance Abuse Treatment, Prevention, and Policy*, *6*, 1-18.

Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics*, *61*, 87–106.

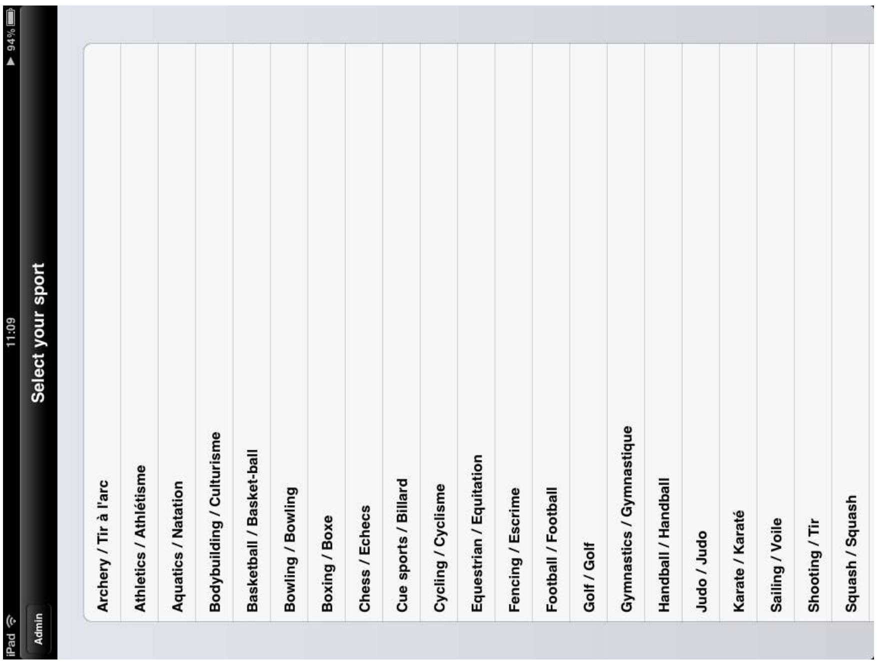

Figure 1: Questionnaire Screen 1: Upper and lower half.

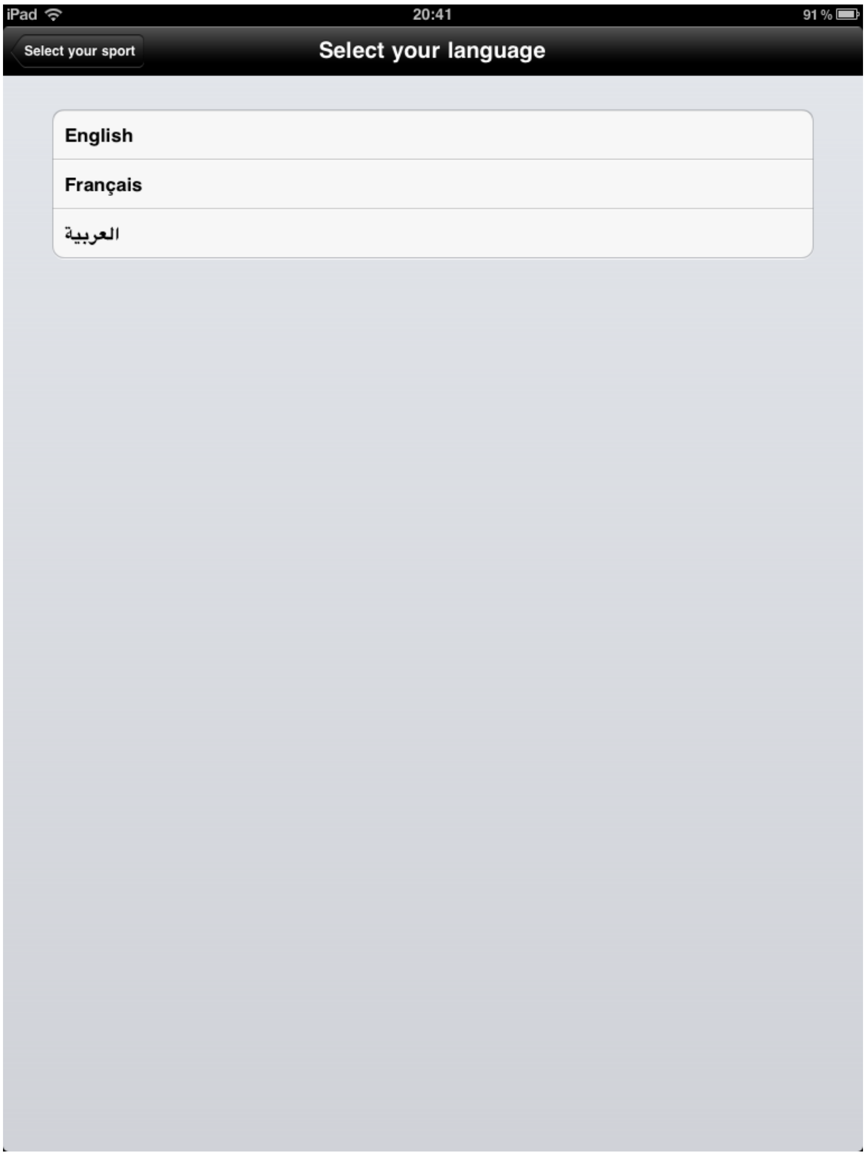

Figure 2: Questionnaire Screen 2.

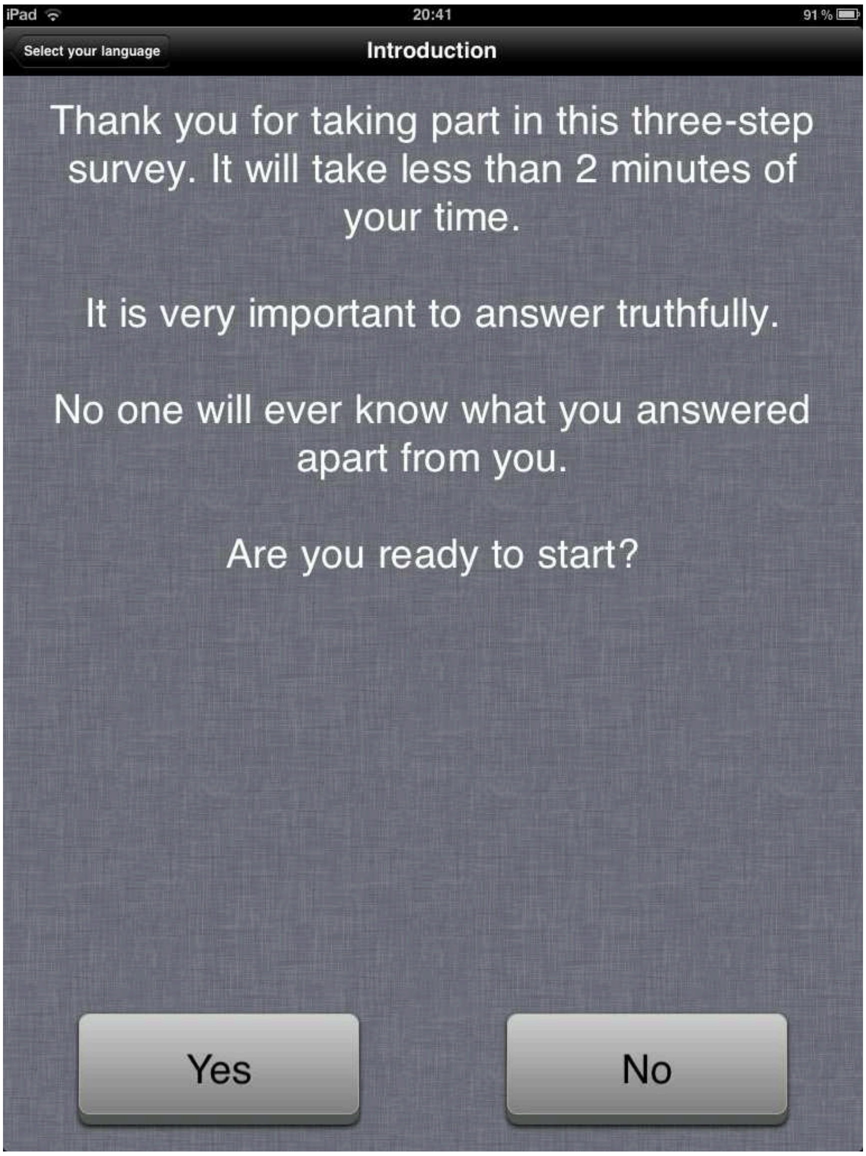

Figure 3: Questionnaire Screen 3.

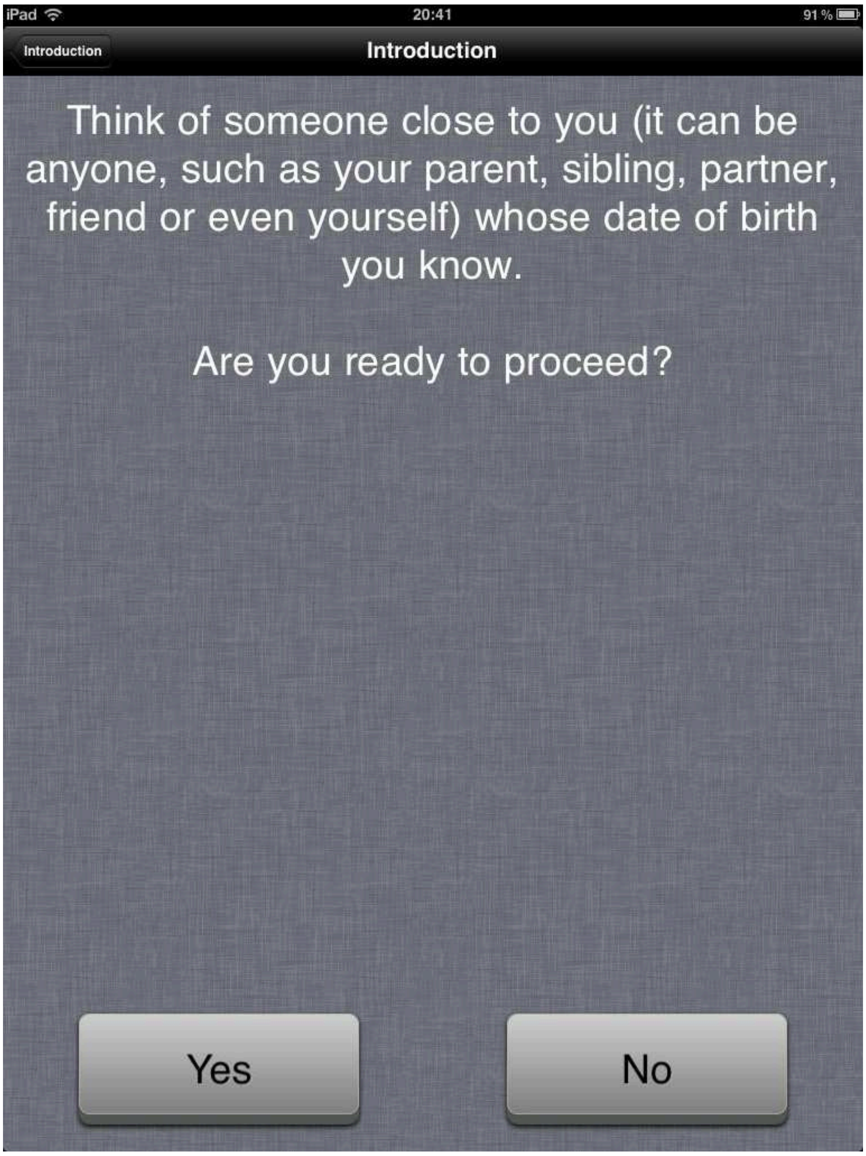

Figure 4: Questionnaire Screen 4.

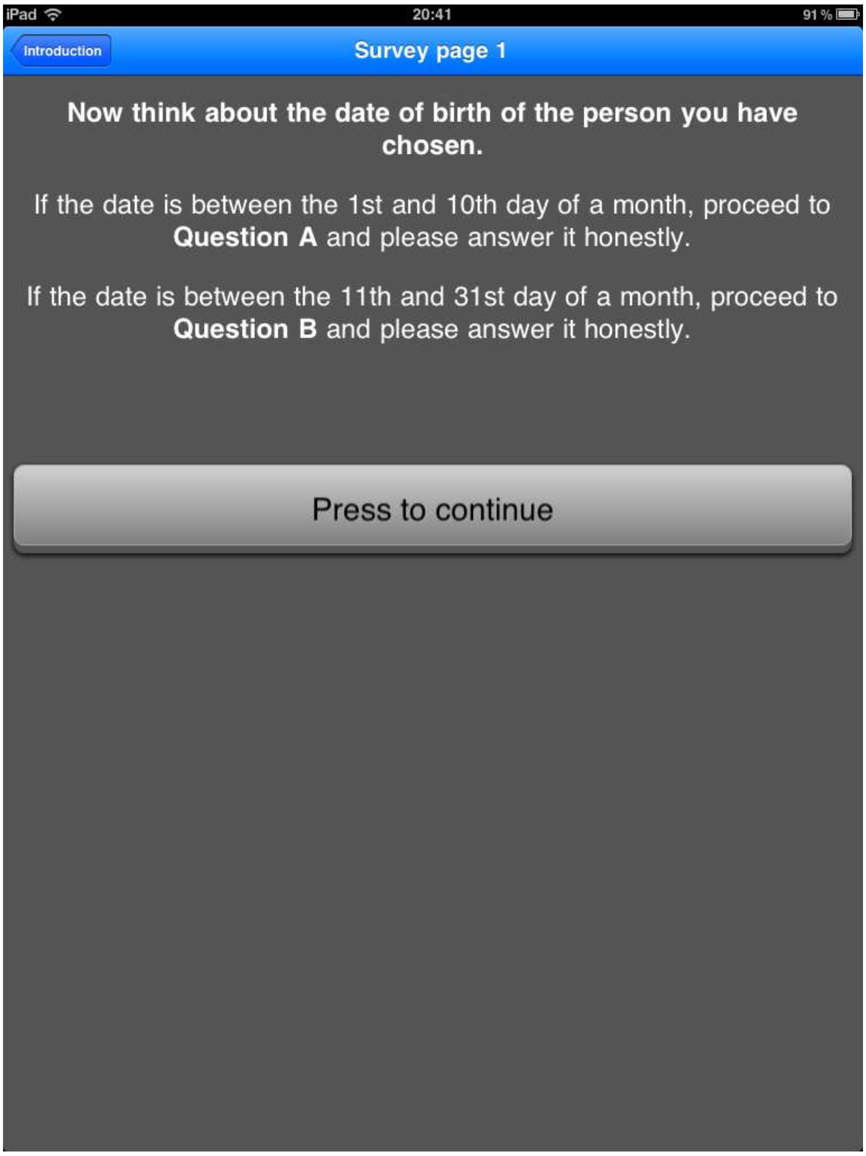

Figure 5: Questionnaire Screen 5.

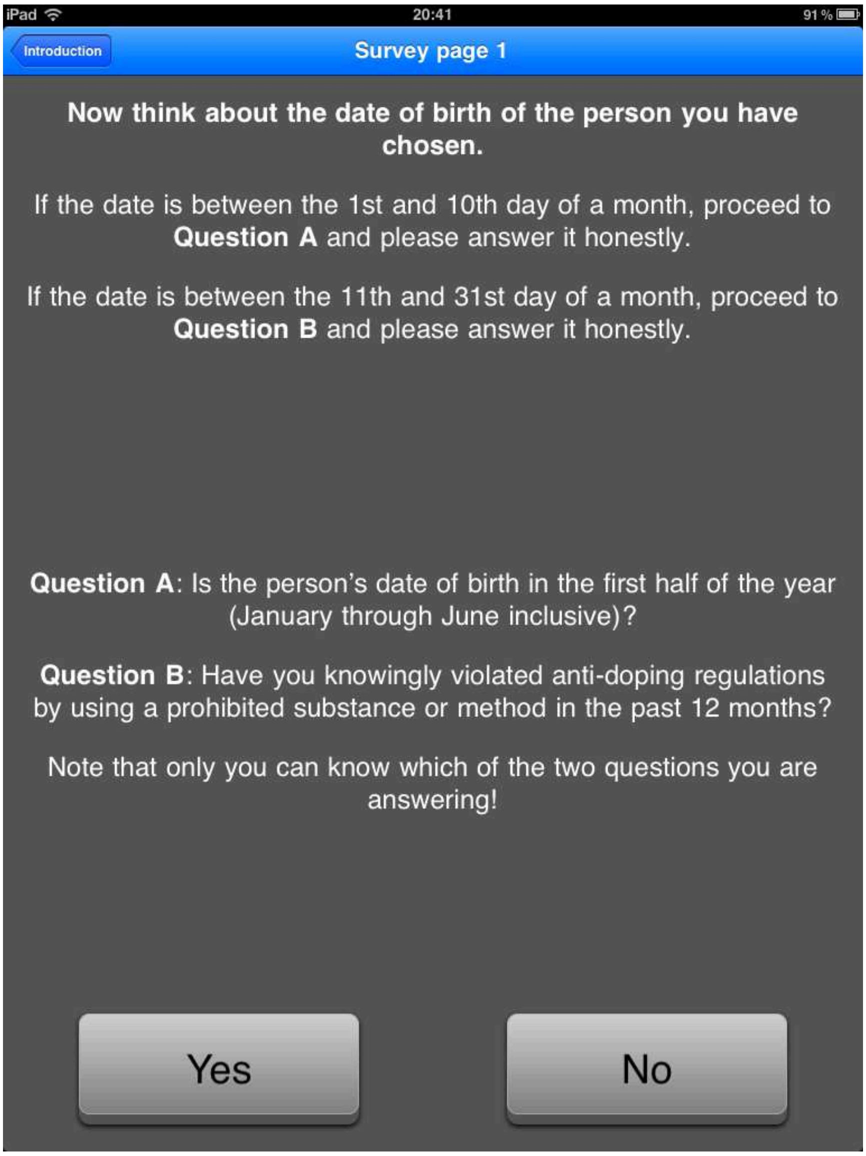

Figure 6: Questionnaire Screen 6.

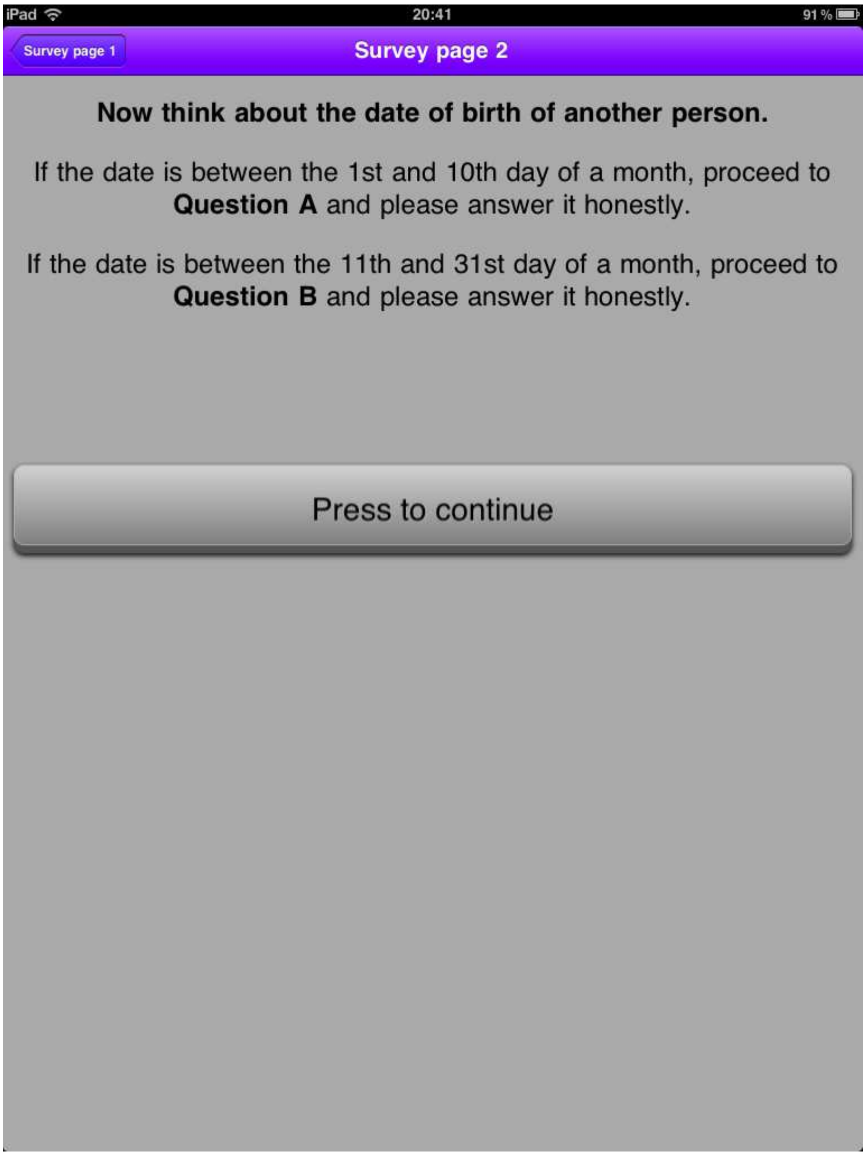

Figure 7: Questionnaire Screen 7.

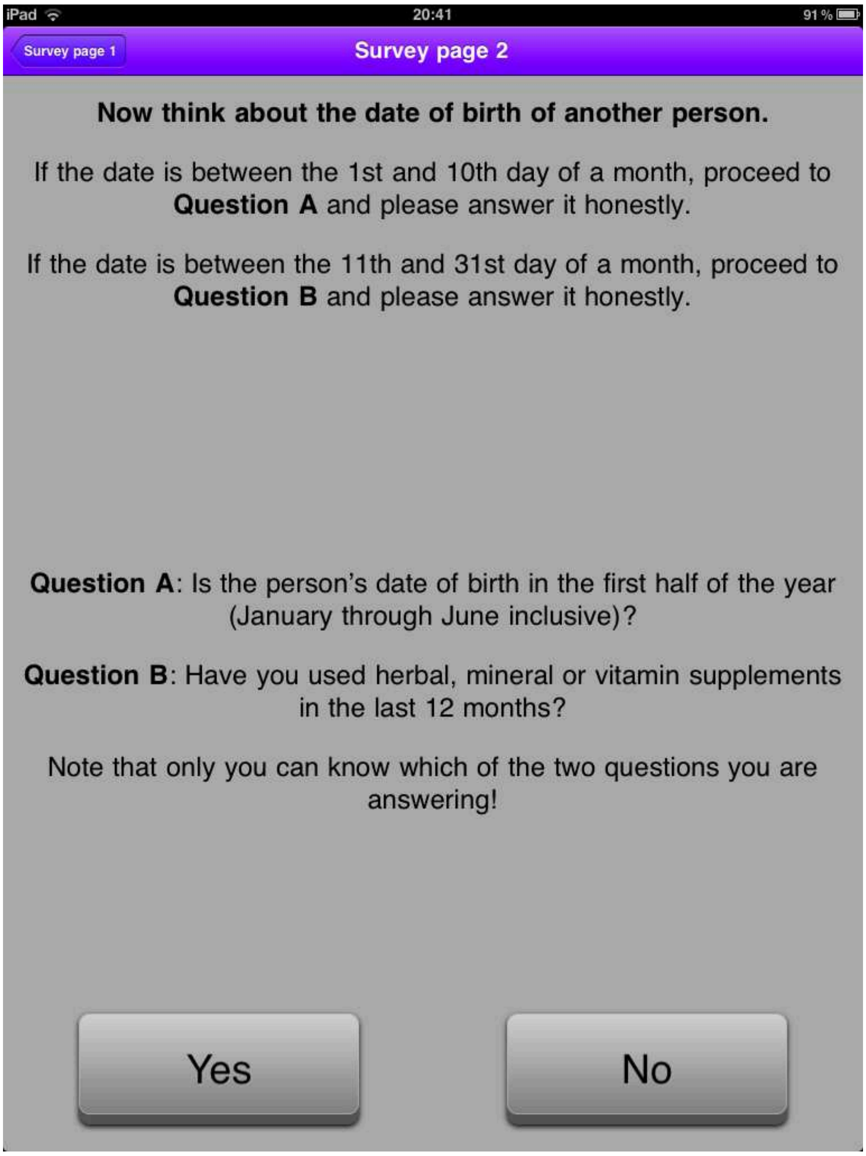

Figure 8: Questionnaire Screen 8.

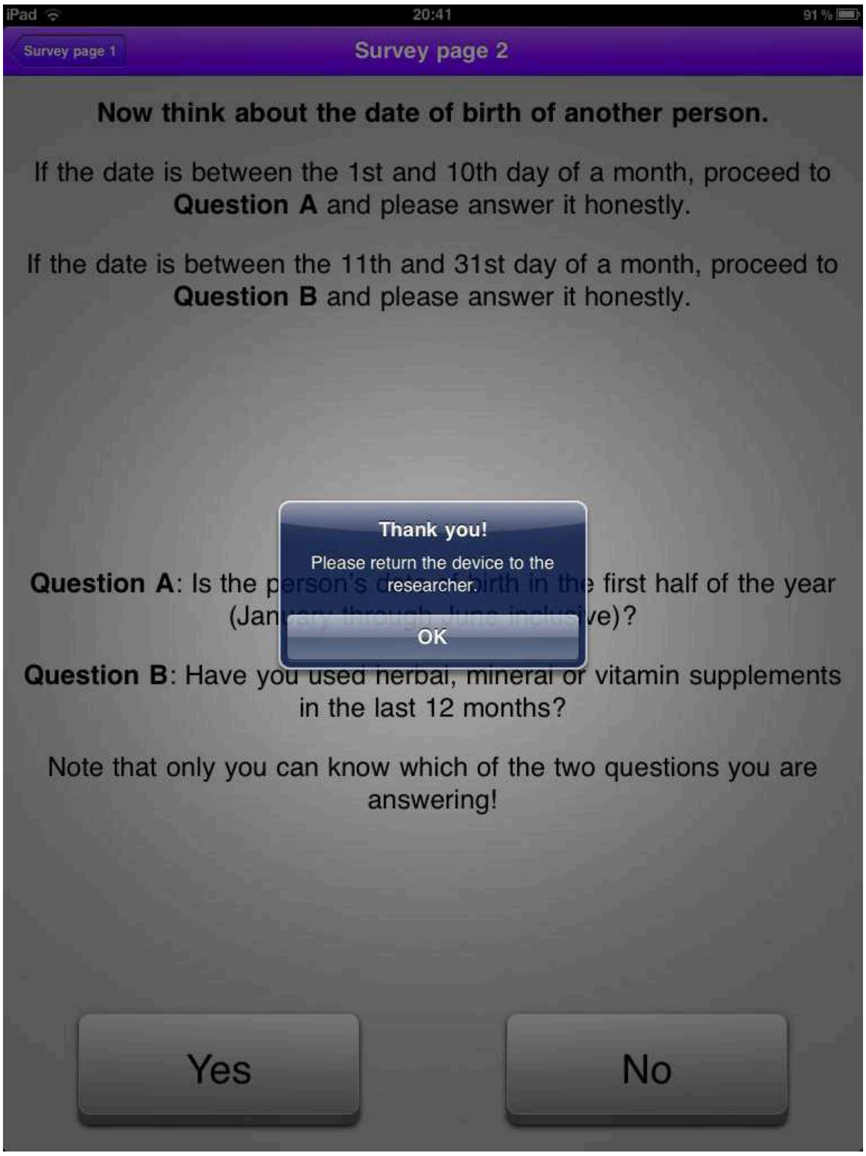

Figure 9: Questionnaire Screen 9.

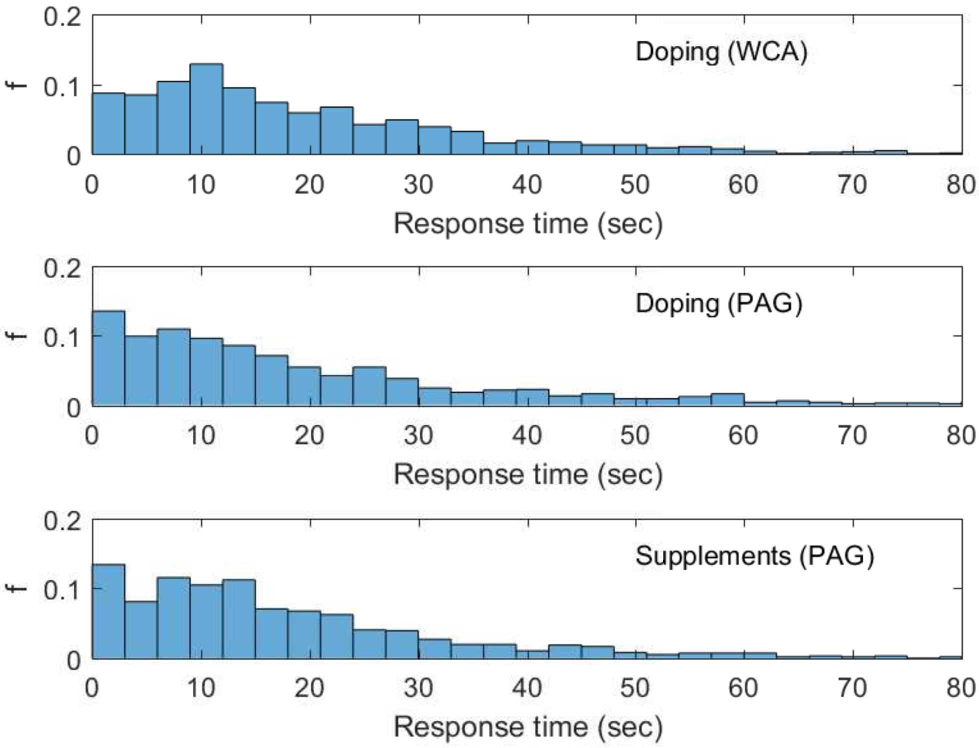

Figure 10: Histograms of the response time data as a function of survey question. The height f of each bar is the relative number of observations (i.e., number of observations in bin divided by total number of observations); thus the sum of the bar heights adds to 1. The width of each bar is 3 sec.

Table 1: Participating Nations and Number of Athletes in WCA.

Afghanistan (1), Albania (1), Algeria (10), American Samoa (2), Angola (2), Anguilla (2), Antigua and Barbuda (2), Argentina (6), Armenia (2), Aruba (2), Australia (41), Austria (4), Azerbaijan (1), Bahamas (17), Bahrain (11), Bangladesh (1), Barbados (4), Belarus (22), Belgium (9), Belize (2), Benin (2), Bermuda(1), Bhutan (1), Bolivia (2), Bosnia and Herzegovina (2), Botswana (3), Brazil (26), British Virgin Islands (1), Brunei (1), Bulgaria (7) Burkina Faso (2), Burundi (2), Cambodia (1), Cameroon (2), Canada (28), Cape Verde (1), Cayman Islands (1), Central African Republic (1), Chad (2), Chile (3), China (54), Colombia (20), Comoros (2), Congo (1), Democratic Republic of the Congo (2), Cook Islands (1), Costa Rica (2), Ivory Coast (2), Croatia (6), Cuba (31), Cyprus (2), Czech Republic (21), Denmark (6), Djibouti (2), Dominica (1), Dominican Republic (4), Ecuador (5), Egypt(5), El Salvador (2), Equatorial Guinea (0), Eritrea (9), Estonia (9), Ethiopia (34), F.S. Micronesia (2), Fiji (1), Finland (13), France (39), French Polynesia (1), Gabon (2), Gambia (2), Germany (65), Ghana (6), Gibraltar (1), Great Britain (59), Greece (12), Grenada (3), Guam (2), Guatemala (2), Guinea (2), Guinea-Bissau (2), Guyana (1), Haiti (3), Honduras (2), Hong Kong (2), Hungary (12), Iceland (2), India (8), Indonesia (2), Iran (7), Iraq (1), Ireland (16), Israel (4), Italy (30), Jamaica (45), Japan (48), Kazakhstan (14), Kenya (47), Kiribati (2), South Korea (53) (Hosts), Kuwait (2), Kyrgyzstan (2), Laos (2), Latvia (13), Lebanon (1), Lesotho (2), Liberia (2), Libya (1), Lithuania (15), Macau (1), Macedonia (1), Madagascar (1), Malawi (2), Malaysia (2), Maldives (2), Mali (2), Malta (2), Marshall Islands (0), Mauritania (2), Mauritius (2), Mexico (10), Moldova (3), Monaco (1), Mongolia (2), Montenegro (2), Morocco (19), Mozambique (2), Myanmar (1), Namibia (2), Nauru (2), Nepal (2), Netherlands (17), New Zealand (8), Nicaragua (2), Niger (2), Nigeria (15), Northern Mariana Islands (2), Norway (13), Oman (1), Pakistan (1), Palau (2), Palestine (1), Panama (2), Papua New Guinea (2), Paraguay (1), Peru (5), Philippines (2), Poland (37), Portugal (25), Puerto Rico (8), Qatar (4), Romania (8), Russia (76), Rwanda (2), Saint Kitts and Nevis (4), Saint Lucia (2), Saint Vincent and the Grenadines (1), Samoa (1), San Marino (2), São Tomó and Príncipe (2), Saudi Arabia (8), Senegal (2), Serbia (9), Seychelles (2), Sierra Leone (2), Singapore (2), Slovakia (8), Slovenia (9), Solomon Islands (2), Somalia (1), South Africa (32), Spain (43), Sri Lanka (2), Sudan (3), Suriname (2), Swaziland (2), Sweden (16), Switzerland (15), Syria (1), Chinese Taipei (7), Tajikistan (2), Tanzania (1), Thailand (6), Timor-Leste (1), Togo (1), Tonga (2), Trinidad and Tobago (16), Tunisia (5), Turkey (20), Turkmenistan (2), Turks and Caicos Islands (0), Tuvalu (2), Uganda (12), Ukraine (55), United Arab Emirates (2), United States (127), Uruguay (2), U.S. Virgin Islands (3), Uzbekistan (7), Vanuatu (2), Venezuela (3), Vietnam (1), Yemen (2), Zambia (3), Zimbabwe(4).

Source: `https://en.m.wikipedia.org/wiki/2011_World_Championships_in_Athletics`

Table 2: Participating Nations and Number of Athletes in PAG.

Algeria (223), Bahrain (171), Comoros (21), Djibouti (38), Egypt (349), Iraq (436), Jordan (240), Kuwait (260), Lebanon (99), Libya (148), Mauritania (43), Morocco (253), Oman (94), Palestine (109), Qatar (370), Saudi Arabia (232), Somalia (22), Sudan (191), Syria (withdrew), Tunisia (218), United Arab Emirates (144), Yemen (46).

Source: `https://en.m.wikipedia.org/wiki/2011_Pan_Arab_Games`

Table 3: Number and percentage of survey respondents in every PAG sport category. The last column gives the percentage for all athletes at the PAG games.

Sport category	Number of survey respondents	Survey respondents (%)	All PAG athletes (%)
Aquatics	11	1.1	3.4
Archery	25	2.6	1.8
Athletics	143	14.8	9.2
Basketball	66	6.8	6.5
Bodybuilding	11	1.1	0.9
Bowling	8	0.8	1.3
Boxing	36	3.7	2.0
Chess	20	2.1	3.4
Cue sports	12	1.2	1.7
Cycling	12	1.2	2.3
Equestrian	4	0.4	2.2
Fencing	7	0.7	3.4
Football	53	5.5	6.8
Goalball	7	0.7	1.9
Golf	12	1.2	1.6
Gymnastics	27	2.8	2.3
Handball	78	8.1	6.6
Judo	37	3.8	3.2
Karate	16	1.7	4.5
Paralympic	4	0.4	4.1
Sailing	34	3.5	2.0
Shooting	63	6.5	6.8
Squash	11	1.1	0.6
Table tennis	53	5.5	4.7
Taekwondo	67	7.0	4.4
Volleyball	97	10.1	7.0
Weightlifting	30	3.1	2.9
Wrestling	21	2.2	2.7

Table 4: Number of athletes N, percentage of yes responses $\hat{\lambda}$, prevalence estimate $\hat{\pi}_S$ (%), standard error of estimate SE (%), and 95% confidence interval CON of estimate as function of the percentage of fast readers deleted from the sensitivity analysis.

	Percentage of fastest readers deleted from sensitivity analysis					
	0%	10%	20%	30%	40%	50%
	Doping (WCA)					
N	1202	1082	962	841	721	601
$\hat{\lambda}$	45.8	41.9	39.4	37.6	37.0	36.6
$\hat{\pi}_S$	43.6	37.8	34.1	31.4	30.5	29.9
SE	2.2	2.2	2.4	2.5	2.7	2.9
CON	39.4-47.9	33.4-42.2	29.5-38.7	26.5-36.3	25.3-35.8	24.1-35.7
	Doping (PAG)					
N	965	869	772	676	579	483
$\hat{\lambda}$	54.7	51.8	49.7	49.4	48.2	46.8
$\hat{\pi}_S$	57.1	52.7	49.6	49.1	47.3	45.2
SE	2.4	2.5	2.7	2.9	3.1	3.4
CON	52.4-61.8	47.7-57.7	44.3-54.9	43.5-54.8	41.2-53.4	38.5-51.9
	Supplements (PAG)					
N	965	869	772	676	579	483
$\hat{\lambda}$	63.4	62.6	60.6	58.9	59.8	59.6
$\hat{\pi}_S$	70.1	68.9	65.9	63.3	64.6	64.4
SE	2.3	2.5	2.6	2.8	3.1	3.3
CON	65.6-74.7	64.1-73.7	60.8-71.1	57.7-68.9	58.6-70.6	57.9-71.0

Table 5: Prevalence estimate and response time measures when the question was presented first or second in the survey. The measures are the number (N) of athletes, percentage of "yes"-responses ($\hat{\lambda}$), prevalence estimate ($\hat{\pi}_S$), standard error (SE) associated with this estimate, mean response time (Mean), standard deviation (SD) of response time, median response time (Median), and the interquartile range (IQR) of the response time distribution.

	Temporal Position	
	First	Second
	Doping (WCA)	
N	599	603
$\hat{\lambda}$	45.9%	45.6%
$\hat{\pi}_S$	43.9%	43.4%
SE	3.1%	3.0%
Mean	22.0 sec	18.3 sec
SD	19.3 sec	16.2 sec
Median	16.5 sec	13.7 sec
IQR	21.9 sec	15.3 sec
	Doping (PAG)	
N	499	466
$\hat{\lambda}$	56.5%	52.8%
$\hat{\pi}_S$	59.8%	54.2%
SE	3.3%	3.5%
Mean	25.1 sec	15.0 sec
SD	22.0 sec	17.0 sec
Median	19.2 sec	10.0 sec
IQR	25.4 sec	14.3 sec
	Supplements (PAG)	
N	466	499
$\hat{\lambda}$	65.2%	61.7%
$\hat{\pi}_S$	72.9%	67.6%
SE	3.3%	3.3%
Mean	23.3 sec	14.4 sec
SD	21.2 sec	13.2 sec
Median	17.5 sec	11.6 sec
IQR	20.8 sec	13.7 sec

Table 6: Response frequencies observed with the SSC method when this method was presented first or second in the survey. The first column depicts the possible number k of "yes" answers to the 5 SSC questions presented in each of the surveys. For example, $k = 3$ means that 3 of the 5 questions were answered with "yes". The second and third columns show the number of athletes responding at each level of k.

	Temporal Position	
k	First	Second
	Doping (WCA)	
0 or 5	78	76
1	169	164
2	211	215
3	95	90
4	50	55
	Doping (PAG)	
0 or 5	40	37
1	134	140
2	139	169
3	97	115
4	45	38
	Supplements (PAG)	
0 or 5	33	43
1	152	118
2	175	167
3	99	83
4	40	44

Table 7: Distortion of the obtained prevalence estimates by random responding. The first column gives the percentage c of respondents assumed to have randomly chosen "yes" or "no" at each event. The remaining columns show the effect of c on the estimated prevalence. Note that $c = 0$ reflects the standard UQM and thus gives the estimated prevalences reported in the main text.

	Prevalence $\hat{\pi}_S$ (%)		
c (%)	Doping (WCA)	Doping (PAG)	Supplements (PAG)
0	43.6	57.1	70.1
10	43.0	57.8	72.3
20	42.1	58.8	75.1
30	41.0	60.1	78.7

Table 8: Distortion of the obtained prevalence estimates by underreporting doping. The first column gives the percentage c of dopers assumed to have responded with "no" to the sensitive question. Note that $c = 0$ reflects the standard UQM and thus gives the estimated prevalences reported in the main text.

	Prevalence $\hat{\pi}_S$ (%)		
c (%)	Doping (WCA)	Doping(PAG)	Supplements (PAG)
0	43.6	57.1	70.1
10	48.6	63.4	77.9
20	54.6	71.3	87.6
30	63.4	81.5	not possible

Table 9: Distortion of the Obtained Prevalence Estimates by Switching from Question B to Question A. Part A: Only dopers switch. Part B: Twice as many dopers as non-dopers switch. Part C: An equal number of dopers and non-dopers switch. The first column gives the percentage g_D of dopers that switch.

	Prevalence $\hat{\pi}_S$ (%)		
	Doping (WCA)	Doping (PAG)	Supplements (PAG)
g_D (%)	Part A: $g_N = 0$ and $g_D > 0$		
0	43.6	57.1	70.1
10	46.0	60.1	73.8
20	48.6	63.4	77.9
30	51.4	67.1	82.5
g_D (%)	Part B: $g_N = g_D/2$		
0	43.6	57.1	70.1
10	44.5	59.0	73.1
20	45.5	61.2	76.6
30	46.7	63.9	80.8
g_D (%)	Part C: $g_N = g_D$		
0	43.6	57.1	70.1
10	43.0	57.8	72.3
20	42.1	58.8	75.1
30	41.0	60.1	78.7

Table 10: Distortion of the Obtained Prevalence Estimates by Switching from Question A to Question B. Part A: Only non-dopers switch. Part B: More non-dopers than dopers switch. Part C: An equal number of dopers and non-dopers switch. The first column gives the percentage g_N of non-dopers that switch.

	Prevalence $\hat{\pi}_S$ (%)		
	Doping (WCA)	Doping (PAG)	Supplements (PAG)
g_N (%)	Part A: $g_N > 0$ and $g_D = 0$		
0	43.6	57.1	70.1
10	45.1	58.1	70.8
20	46.4	59.1	71.5
30	47.6	60.0	72.2
g_N (%)	Part B: $g_N = 2 \cdot g_D$		
0	43.6	57.1	70.1
10	44.5	57.4	70.0
20	45.3	57.7	69.9
30	46.0	58.0	69.8
g_N (%)	Part C: $g_N = g_D$		
0	43.6	57.1	70.1
10	44.0	56.7	69.1
20	44.3	56.4	68.3
30	44.5	56.1	67.5

Table 11: Distortion of the obtained prevalence estimates by ignoring the instructions and jumping directly to Question B. The first column gives the percentage c of athletes assumed to have jumped to Question B. Note that $c = 0$ reflects the standard UQM and thus gives the estimated prevalences reported in the main text.

	Prevalence $\hat{\pi}_S$ (%)		
c (%)	Doping (WCA)	Doping (PAG)	Supplements (PAG)
0	43.6	57.1	70.1
10	44.0	56.7	69.1
20	44.3	56.4	68.3
30	44.5	56.1	67.5

Table 12: Distortion of the obtained prevalence estimates by automatic responding regardless of the question received. Part A: Automatic "yes" responding. The first column gives the percentage a of athletes assumed to have automatically responded with "yes". Part B: Automatic "no" responding. The first column gives the percentage a of athletes assumed to have automatically responded with "no". In both parts $a = 0$ reflects the standard UQM estimate and thus gives the estimated prevalences reported in the main text.

	Prevalence $\hat{\pi}_S$ (%)		
	Doping (WCA)	Doping (PAG)	Supplements (PAG)
a (%)	Part A: Automatic "yes" responding		
0	43.6	57.1	70.1
10	34.7	49.5	64.0
20	23.4	40.1	56.4
30	8.9	27.9	46.6
a (%)	Part B: Automatic "no" responding		
0	43.6	57.1	70.1
10	51.3	66.2	80.7
20	60.9	77.6	93.9
30	73.1	92.2	NaN

Note – NaN indicates that this scenario is not possible for $a = 30\%$.

Table 13: Possible scenarios of noncompliance

False "no"			
By non-dopers		By dopers	
Receive Question A	Receive Question B	Receive Question A	Receive Question B
Falsely answer "no" when the answer should have been "yes", due to either (a) a mistake (e.g., finger error), or (b) automatic "no-saying"	(Not possible, since "no" is the correct answer to B by definition)	Falsely answer "no" when the answer should have been "yes", due to either (a) a mistake, (b) automatic "no-saying", or (c) cheating	Falsely answer "no" due to (a) outright lying, (b) switching to Question A and answering it "no", or (c) automatic "no-saying"
False "yes"			
By non-dopers		By dopers	
Receive Question A	Receive Question B	Receive Question A	Receive Question B
Falsely answer "yes" when the answer should have been "no", due to either (a) a mistake (e.g., finger error), or (b) automatic "yes-saying"	Falsely answer "yes" due to (a) false self-incrimination, (b) switching to Question A and answering it "yes", or (c) automatic "yes-saying"	Falsely answer "yes" when the answer should have been "no", due to either (a) a mistake, (b) automatic "yes-saying", or (c) a desire to reveal doping	(Not possible, since "yes" is the correct answer to B by definition)

Table 14: Summary of Potential Violations of the Standard UQM. For each form of noncompliance considered in this table we assume that 30% of athletes engaged in the particular non-compliant behavior in question (the highest level of noncompliance considered in each of our tables above). Symbols are defined as follows: "−" and "+" indicate that our standard UQM estimate would underestimate and overestimate, respectively, the true prevalence within our sample by more than 3%, whereas "±" indicates a maximum error of ±3% between our standard UQM estimate and the true prevalence within our sample.

			Bias		
Scenario	Equation	Table	Doping (WCA)	Doping (PAG)	Supplements
Random Responding	2	7	±	−	−
Underreporting Doping	3	8	−	−	−
Switch from B to A (dopers)	6	9A	−	−	−
Switch from B to A (mostly dopers)	6	9B	−	−	−
Switch from B to A (equal case)	6	9C	±	−	−
Switch from A to B (non-dopers)	8	10A	−	−	±
Switch from A to B (mostly non-dopers)	8	10B	±	±	±
Switch from A to B (equal case)	8	10C	±	±	±
Responding to Question A	2	7	±	−	−
Responding to Question B	11	11	±	±	±
Automatic "yes" Responding	14	12	+	+	+
Automatic "no" Responding	17	12	−	−	−
Cheating with "no" Responding	25	—	−	−	−

Note — The conclusion for the scenario "Cheating with 'no' responding" is based on analytical analysis rather than on numerical results (Section 4.7).