

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

Doping in Two Elite Athletics Competitions Assessed by Randomized-Response Surveys

Rolf Ulrich¹

Harrison G. Pope, Jr.^{2*}

Léa Cléret³

Andrea Petróczi^{4,5}

Tamás Nepusz^{4,6}

Jay Schaffer⁷

Gen Kanayama²

R. Dawn Comstock⁸

Perikles Simon⁹

¹ Department of Psychology, University of Tübingen, D-72076 Tübingen, Germany

² Biological Psychiatry Laboratory, McLean Hospital, Belmont Massachusetts, USA 02478 and Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA.

³ Honorary Research Fellow, Sport and Exercise Science, Swansea University, Swansea, UK.

⁴ School of Life Sciences, Pharmacy and Chemistry, Kingston University, Kingston-upon-Thames, Surrey, KT1 2EE, United Kingdom

⁵ Department of Psychology, The University of Sheffield, Sheffield, S10 2TP, United Kingdom

⁶ Faculty of Logistics, Molde University College, Molde, Norway

⁷ Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, Colorado, USA

⁸ Pediatric Injury Prevention, Education, and Research Program, Colorado School of Public Health, Aurora, CO, USA

⁹ Department of Sports Medicine, Rehabilitation and Disease Prevention, Johannes Gutenberg University Mainz, D-55131 Mainz, Germany

*To whom correspondence should be addressed. E-mail: hpoppe@mclean.harvard.edu

Abstract

Background Doping in sports compromises fair play and endangers health. To deter doping among elite athletes, the World Anti-Doping Agency (WADA) oversees testing of several hundred thousand athletic blood and urine samples annually, of which 1-2% test positive. Measures using the Athlete Biological Passport suggest a higher mean prevalence of about 14% positive tests. Biological testing, however, likely fails to detect many cutting-edge doping techniques, and thus the true prevalence of doping remains unknown.

Methods We surveyed 2167 athletes at two sporting events: the 13th International Association of Athletics Federations World Championships in Athletics (WCA) in Daegu, South Korea in August 2011 and the 12th Quadrennial Pan-Arab Games (PAG) in Doha, Qatar in December 2011. To estimate the prevalence of doping, we utilized a “randomized response technique,” which guarantees anonymity for individuals when answering a sensitive question. We also administered a control question at PAG assessing past-year use of supplements.

Results The estimated prevalence of past-year doping was 43.6% (95% confidence interval 39.4-47.9%) at WCA and 57.1% (52.4-61.8%) at PAG. The estimated prevalence of past-year supplement use at PAG was 70.1% (65.6-74.7%). Sensitivity analyses, assessing the robustness of these estimates under numerous hypothetical scenarios of intentional or unintentional noncompliance by respondents, suggested that we were unlikely to have overestimated the true prevalence of doping.

Conclusions Doping appears remarkably widespread among elite athletes, and remains largely unchecked despite current biological testing. The survey technique presented here will allow future investigators to generate continued reference estimates of the prevalence of doping.

Key Points

- Although elite athletes routinely receive biological tests to detect prohibited doping, these tests likely fail to detect many cutting-edge doping techniques, and thus the true prevalence of doping remains unknown.
- We utilized a “randomized response technique” – a method that guarantees anonymity for individuals when answering a sensitive question – to estimate the prevalence of past-year doping at two major international athletic events: the 13th International Association of Athletics Federations World Championships in Athletics (WCA) in Daegu, South Korea and the 12th Quadrennial Pan-Arab Games (PAG) in Doha, Qatar, both held in 2011.
- After performing numerous sensitivity analyses, assessing the robustness of our estimates under various hypothetical scenarios of intentional or unintentional noncompliance by respondents, we found that the prevalence of past-year doping was at least 30% at WCA and 45% at PAG.
- These findings suggest that biological testing greatly underestimates the true prevalence of doping in elite athletics, and indicate the need for future studies of the prevalence of doping in athletics using randomized response techniques.

1 Introduction

Recent decades have seen increasingly sophisticated biological testing programs to deter doping among athletes, especially at elite international competitions [1–3]. Among Olympic-level athletes tested between 1987 and 2013, the percentage of positive test results has ranged from 0.96% to 2.45% [1]. However, with sophisticated modern doping schemes [4–8], many athletes may beat the tests. Given the numerous recent highly publicized doping scandals in major sports [9–11], one might guess that the proportion of such undetected cheats is high. Nevertheless, the true rates of both false-negative and false-positive cases among tested athletes remain unknown, while subject to much continuing speculation and debate [12–15]. Several recent commentaries have suggested technical, human, political, and financial factors that may have contributed to flawed results from current biological testing techniques [4,16,17]. Unfortunately, however, one cannot estimate the rates of such false test results without an estimate of the true underlying prevalence of doping [12,18].

To pursue this issue, we assessed the prevalence of doping at two major international athletic competitions using a novel method, the so-called “randomized response technique (RRT).” This technique allows investigators to pose sensitive questions to respondents in a manner that visibly guarantees the respondent’s anonymity – thus encouraging truthful responses. The basic idea of the technique and its variants is to include a random element that masks each individual’s answer to the sensitive question, thereby encouraging honest reports (for reviews see [19,20]). An extensive meta-analysis of RRT research [21] has shown that these survey techniques indeed result in more valid data than conventional question-and-answer methods.

The RRT was originally developed by Warner in 1965 [22] to estimate the proportion π_s of “yes” answers to sensitive questions on issues where a respondent might be reluctant to disclose the true answer for fear of loss of confidentiality or other reasons. In Warner’s model, the respondent receives one of two questions about a sensitive issue. For example, Question A might be, “*have you ever used illicit drugs?*” while Question B would be the reverse question, “*have you never used illicit drugs?*” There is a known probability p that the

respondent will receive Question A, and the complementary probability $1 - p$ that the respondent will receive Question B. A random element (e.g., the throw of a die seen only by the respondent) determines which of the two questions the respondent receives. Thus, when the investigators obtain a “yes” or “no” answer from a given respondent, they cannot know the status of that specific respondent on the sensitive item, since they cannot know which of the two questions that particular individual has received and answered. However, given a large survey population, and knowing the value of p and the total number of “yes” answers, the investigators can estimate the prevalence of the sensitive item in the overall population.

However, one limitation of Warner’s original technique is that both questions involve the sensitive topic. Thus, some respondents may believe that there is a trick that enables the investigators to figure out their real status on the sensitive item. To address this problem and make Warner’s procedure psychologically more acceptable, Greenberg, Abul-Ela, Simmons, and Horvitz in 1969 [23] proposed the unrelated question model (UQM; see Figure 1).

Although the basic procedure of the UQM is analogous to Warner’s technique, Question B is replaced by a neutral question, such as “*think of someone close to you whose birth date you know, and answer ‘yes’ if that person was born during the first half of the year.*” In this example, the probability π_N of a “yes” response to Question B will be approximately $\pi_N = 0.5$. Note that although birth rates vary slightly across months [24], it can be shown that UQM estimates are rather robust when the true probability π_N deviates from the assumed probability [25]. Thus, the UQM will be similarly robust if some respondents misinterpret a question (e.g., “*were you born during the first half of the year?*”) and respond incorrectly.

In addition to being psychologically more acceptable than Warner’s original method, the UQM also has more favorable statistical properties (see [25,26]). With the UQM, the maximum likelihood estimator of π_S is computed from the observed proportion λ of total “yes” responses (see [23], p. 533) as

$$\hat{\pi}_S = \frac{\lambda - (1-p) \cdot \pi_N}{p}. \quad (1)$$

With a sample of n respondents, the standard error of the estimate $\hat{\pi}_S$ is computed as

$$SE = \sqrt{\frac{\lambda \cdot (1-\lambda)}{n \cdot p^2}}. \quad (2)$$

With adequate sample size, this technique yields high statistical power [25].

Randomized response techniques such as the UQM have been used in several prior studies of athletes in Germany [27–33], which all revealed a much higher prevalence of doping among athletes than that obtained with conventional survey methods that lacked a guarantee of anonymity. The first such study RRT study, to our knowledge, was performed by Musch and Plessner in 2002 [27]. In an online survey, these authors employed a forced-answer version of RRT and surveyed 467 German athletes; this study reported a 42% prevalence of doping for the athletes during their careers. In another online survey in 2007, Pitsch, Emrich and Klein [28] also employed the forced-answer version of RRT to assess doping among a group of 448 German top-level athletes. These researchers reported that between 26% and 48% of this group admitted to having used banned substances or methods during their athletic careers. In 2012, using a much larger sample, the same investigators replicated their previous finding and concluded that between 10% and 35% of German athletes use doping in a given season. In a recent RRT online study, Frenger, Pitsch and Emrich estimated the doping prevalence among about 2000 recreational and amateur athletes [33]. This study reported a lower limit of about 4% for the use of prohibited substances (i.e., approximately 4% for both lifetime use and for use during the last season). In another RRT study using the UQM, Striegel, Ulrich and Simon [29] assessed intentional doping behavior among 480 junior athletes in Germany. They estimated a lifetime prevalence of 6.8% for this group of athletes. Twelve-month prevalence estimates between 4% and 12% were observed in two German triathlon events among 2017 athletes based on two different RRT methods (a forced-choice RRT version and UQM) [32]. A similar RRT survey using UQM at three locations in Germany revealed a 12-month prevalence of 13% among a sample of 2997 triathletes [31].

Based on the results of RRT studies, together with other sources (e.g., blood doping markers, and questionnaires), de Hon et al. [1] concluded, in a review of doping in elite

sports, that about 14-39% of elite athletes have employed doping. These authors suggest that RRT studies currently yield the most accurate estimates of doping in sports, and they suggest that this approach should be used routinely to assess the prevalence of doping. However, the authors note that “true doping prevalence studies have been scarce in elite sport so far” (p. 57).

2 Methods

2.1 Study design and participants

We utilized the UQM method to estimate the prevalence of past-year doping at two elite international athletic competitions: the 13th International Association of Athletics Federations World Championships in Athletics (WCA) in Daegu, South Korea and the 12th Quadrennial Pan-Arab Games (PAG) in Doha, Qatar, both held in 2011. We chose these two events because they conveniently occurred a short time apart, allowing us to refine our methods at PAG based on experience from WCA. Thus, the study was not specifically designed to compare doping rates at these two events.

Six data collectors, selected because they collectively spoke 10 languages, approached athletes at WCA and invited them to perform the survey on tablet computers. The survey was available in 21 languages (English, Spanish, Russian, French, Korean, Arabic, Portuguese, Japanese, Polish, German, Chinese, Czech, Swahili, Turkish, Greek, Norwegian, Romanian, Slovakian, Estonian, Croatian and Slovenian). The text of the survey instrument in all languages was translated by the World Anti-Doping Agency translation services, and then back-translated by staffs of the National Anti-Doping Agencies most appropriate to each language in order to maximize accuracy. At PAG, with its less geographically diverse athlete population, the survey was available in English, Arabic and French, and was administered by 10 data collectors who collectively spoke five languages. We obtained excellent cooperation: of the 1841 registered athletes at WCA, we approached 1290, of whom 1203 (93.3%) agreed to participate. Of 3346 athletes accredited at PAG, we approached 1030, of whom 965 (93.4%) agreed to participate. At both games, the number of

athletes approached was limited simply by the availability of data collectors and tablet computers; the athletes approached appeared representative of the entire population registered. Additional information about these samples appears in the appendix, section 1.

2.2 Procedure

The tablet computers displayed the survey questions and recorded both the athletes' responses ("yes" or "no") and their response times. The latter measure enabled us to identify athletes who responded very quickly and hence perhaps carelessly. The first screen seen by the athletes (used in PAG but not at WCA) asked them to indicate the sport in which they were competing; this screen also included a prominent red button that athletes could press if they wished to decline participation. The next screen offered athletes a choice of languages as described above. A detailed presentation of these and all subsequent screens, together with an explanation of the minor differences between the instrument administered at WCA and that used at PAG, is provided in section 2.1 of the appendix.

The next screen invited the athlete to proceed with the study; this screen (which now appeared in the athlete's chosen language) included a prominent "no" button offering athletes the option to decline participation (Figure 2a). Additionally, if athletes reached a later screen and showed unwillingness to participate, they were directed back to this "decline" option – thus effectively ensuring a consent process. The subsequent screens utilized the UQM to estimate the prevalence of past-year doping while guaranteeing the athlete's anonymity. Specifically, participants were asked to think of a person close to them (e.g., a parent, sibling, or spouse) whose date of birth they knew (Figure 2b). Then, if this date fell between the 1st and 10th day of the month, the participant was asked to answer a non-sensitive question, Question A: "*Is the person's date of birth in the first half of the year (January through June inclusive)?*" However, if the chosen person's birth date fell between the 11st and 31th day of the month, the participant was asked to answer the sensitive Question B: "*Have you knowingly violated anti-doping regulations by using a prohibited substance or method in the last 12 months?*" (Figures 2c-2d).

We then analyzed the athletes' responses using equations (1) and (2) above, using values of $p = (365.25 - 12.10)/365.25 = 0.671$ and $\pi_N = 0.5$, to calculate estimates and their 95% confidence intervals for the prevalence of past-year doping among athletes at each of the two events. Full methodological details of this procedure are provided in the appendix, sections 1-2.

As a partial test of the validity of this methodology, we added a control question in the survey administered at PAG: "*Have you used herbal, mineral, or vitamin supplements in the past 12 months?*" This question was administered in a manner identical to the doping question (e.g., an introductory "choose a person" screen, followed by Questions A and B, where "B" was the "supplements" question), with the order of doping and control questions counterbalanced across respondents. We then estimated the prevalence of past-year supplement use by the formula just described.

Finally, we performed a series of sensitivity analyses to assess the robustness of our estimates under numerous hypothetical assumptions of noncompliance or carelessness among the respondents, as detailed below.

3 Results

3.1 Primary analysis

We obtained strikingly high estimates of the prevalence of past-year doping at both events: 43.6% (95% confidence interval 39.4-47.9%) at WCA and 57.1% (52.4-61.8%) at PAG. The estimated prevalence of past-year supplement use at PAG was 70.1% (65.6-74.7%). These estimates are markedly greater than the results obtained by biological testing at the two events. Specifically, at WCA, 440 athletes received biological testing, and only 2 (0.5%) were found positive. At PAG, 670 athletes were tested, of whom 24 (3.6%) were positive. Notably, the prevalence of positive analytical findings at PAG was significantly greater than that at WCA ($p < 0.001$, two-tailed; bootstrapping the statistics $D = P_{PAG} - P_{WCA}$, where P_{PAG} and P_{WCA} are the resampled prevalence rates for PAG and WCA, respectively; N

Doping in elite sports assessed by randomized-response surveys

= 100,000 bootstrap samples; the null hypothesis is $\mu_D = 0$). Further details are provided in the appendix, section 3 (basic results).

3.2 Analyses using response time

We noted that some athletes showed very rapid response times, suggesting possible carelessness or failure to adequately read the instructions. Interestingly, at both athletic events, fast responders yielded even higher prevalence estimates than normal responders, possibly as an artifact of hasty responding. Specifically, athletes had to touch “yes” twice on the opening tablet screens to reach the beginning of the survey (appendix section 2 and Figure 2), and thus some hasty athletes might have continued to mechanically respond “yes” on successive screens. Such behavioral rigidity has been documented in prior psychological studies [34]. Accordingly, we performed sensitivity analyses in which we deleted 0%, 10%, 20%, 30%, 40%, or 50% of the fastest responders (see details in appendix, section 3). The prevalence estimates decreased upon deleting the 10-20% of fastest responders, but stabilized thereafter (Figure 3). Looking at the results obtained by deleting 30% or more of the fastest responders, we obtained estimates of 30-31% for past year doping at WCA and 45-49% for PAG.

4 Discussion

4.1 Summary

We used a randomized response technique, a survey method designed to guarantee anonymity to respondents, to assess the prevalence of past-year doping among 2167 elite athletes at two major international competitions. With this method, we found that the prevalence of past-year doping at both events was many times greater than the 1-2% prevalence typically estimated from biological testing, or even the 14% figure estimated from the Athlete Biological Passport [8] (which does not quantify all forms of doping, in contrast to the present RRT survey). Our findings augment the findings of prior studies using the randomized response technique in more local athletic populations [27–33], which have also found doping to be much more prevalent than indicated by conventional survey methods or

by biological testing. These results suggest that doping in elite athletics may be much more prevalent than has been reported before.

4.2 Limitations

The current findings may be limited by the possibility of various types of noncompliance among the respondents. To address this issue, we considered the potential effects of three plausible forms of respondent noncompliance. First, refusal to participate might have introduced selection bias, although any such bias was likely modest, since of the athletes approached, only 6.7% refused participation. Moreover, doping athletes likely refused participation more often than non-doping athletes, thus potentially causing us to underestimate the true prevalence of doping. Second, doping athletes directed to Question B might have lied and answered a self-protective “no” despite the assurance of anonymity. This behavior also would have led us to underestimate the true prevalence (appendix, section 4.2). Third, some dopers, upon receiving Question B, might have surreptitiously reverted to non-sensitive Question A (e.g., they might have retroactively chosen a different “close person” born within the first ten days of the month). This phenomenon again would have biased our estimates downwards, since only 50% of these noncompliant dopers would consequently have answered “yes” (appendix, section 4.3). Thus each of these three possible forms of noncompliance would have caused us to underestimate the true prevalence of doping.

We next considered three other seemingly less plausible scenarios. First, some non-dopers receiving Question A might have reverted to Question B, even though they (unlike dopers) would have had no obvious motivation to do so. Even assuming such unlikely behavior, our estimates would remain little changed. For example, if 20% of dopers and 20% of non-dopers receiving question A had inappropriately reverted to B, our estimates would be biased by less than 2% (appendix, section 4.4). Second, some non-dopers might have falsely claimed to be dopers. However, there is little plausible motivation for such behavior, and empirical evidence suggests that such false self-incrimination is very rare [35]. Third, some athletes, regardless of doping status, might have randomly pressed “yes” or “no”

Doping in elite sports assessed by randomized-response surveys

without actually considering the questions, due to haste or poor reading ability. Such behavior would drive the estimates towards 50%, and again would have little effect on our results. For example, if 20% of athletes had answered randomly, our prevalence estimates would still be biased by less than 2% (appendix, section 4.1). Fourth, we also considered the effect of automatic or hasty responding with "yes" or "no". Although automatic yes-responding and automatic-no responding would substantially overestimate and underestimate, respectively, the prevalence of doping, automatic responders were almost certainly excluded by our exercise of deleting 30% of the fastest responders (see section 3.2 above, and also sections 4.6 and 4.7 in the appendix). Thus our prevalence estimates after deleting the 30% of fastest responders are likely to present a lower bound of the doping prevalence at each event.

Finally, we considered that some athletes using prohibited substances might have erroneously believed that they were not doping, or conversely might have believed that they were doping despite actually complying with doping regulations. However, such errors would seem rare among world-class athletes, who are typically well educated regarding doping rules, and thus unlikely to misinterpret a question explicitly asking about having "knowingly violated anti-doping regulations." Moreover, any such false-negative and false-positive misinterpretations would have biased the results in opposing directions (appendix, section 4.9). Nevertheless, we acknowledge the limitations of utilizing a single "doping" question – chosen to maximize simplicity and participant compliance in this study – and would encourage more detailed questioning in subsequent studies.

Also in our attempt to maximize simplicity and participant compliance, we omitted demographic questions such as age, gender, country of origin, or years of elite competition. Specifically, we considered that some athletes presented with such questions might refuse participation, fearing that their anonymity might be compromised (even though, in reality, the RRT method ensured that we could not know their doping status even if we knew their precise identities). Thus, we could not use these demographic measures to assess the representativeness of the sample of athletes who chose to participate at each event.

Doping in elite sports assessed by randomized-response surveys

However, our data on the distribution of languages chosen by athletes at WCA and the distribution of sports/events endorsed by athletes at PAG suggest that our samples were reasonably representative (see appendix, section 3 and Table 3).

In summary, even assuming conservatively that (a) none of the first three above-mentioned plausible forms of noncompliance caused any underestimate of the doping prevalence, despite their mutual risk of doing so; and (b) the entire surfeit of “yes” responses among the fast responders was artifactual, and did not reflect any surfeit of genuine dopers; we would estimate a 30-31% prevalence of past-year doping among athletes at WCA and 45-49% at PAG. The reasons for the higher PAG estimate remain speculative, but notably, adverse analytical findings on biological testing were also higher at PAG, as noted above.

Our control question at PAG, assessing past-year supplement use, yielded an estimate of 64-65% after deleting fast responders – a finding consistent with the results of a systematic review and meta-analysis on supplement use by elite athletes, which reports an average prevalence of dietary supplement use of 69% (95% confidence interval, 60-78%) for male athletes and 71% (62-79%) for female athletes [36]. Because of the heterogeneity of athlete populations, one must be cautious in comparing our PAG results with those from this meta-analysis. Nevertheless, the consistency of our results with these meta-analytic results offers some support for the validity of our survey methodology.

A comprehensive analysis of possible noncompliance scenarios, detailed in the appendix and summarized in section 4.10, finds no combination of plausible noncompliant response behaviors likely to have caused our findings to be overestimates. By contrast, there are numerous reasons to suspect that we may well have underestimated the true prevalence of doping among these athletes.

5 Conclusion

Our results, combined with the results of previous recent studies [27–33], support the feasibility of using a randomized response technique to estimate the prevalence of doping in elite athletes. With this technique, we found that the prevalence of doping appears to be

much higher than that indicated by biological testing. Given these provocative findings, we would urge continued use and refinement of this methodology to estimate the prevalence of doping in future sports events.

Data availability

The data that support the main conclusions of this study are available in the online repository of Sports Medicine.

References

1. de Hon O, Kuipers H, van Bottenburg M. Prevalence of doping use in elite sports: a review of numbers and methods. *Sport. Med.* 2014;45:57–69.
2. Dirix A. The doping problem at the Tokyo and Mexico City Olympic Games. *J. Sports Med. Phys. Fitness.* 1966;6:183–6.
3. Scarpino V, Garattini S, La Vecchia C, Silvestrini G, Rossi Bernardi L, Tuccimmi G, et al. Evaluation of prevalence of “doping” among Italian athletes. *Lancet.* 1990;336:1048–50.
4. Butch AW, Lombardo JA, Bowers LD, Chu J, Cowan DA. The quest for clean competition in sports: are the testers catching the dopers? *Clin. Chem.* 2011/02/19. 2011;57:943–7.
5. Thomas A, Kohler M, Schänzer W, Delahaut P, Thevis M. Determination of IGF-1 and IGF-2, their degradation products and synthetic analogues in urine by LC-MS/MS. *Analyst.* 2010/12/16. 2011;136:1003–12.
6. Beiter T, Zimmermann M, Fragasso A, Hudemann J, Niess AM, Bitzer M, et al. Direct and long-term detection of gene doping in conventional blood samples. *Gene Ther.* 2011;18:225–31.
7. Ashenden M, Gough CE, Garnham A, Gore CJ, Sharpe K. Current markers of the Athlete Blood Passport do not flag microdose EPO doping. *Eur. J. Appl. Physiol.* 2011;111:2307–14.
8. Sottas PE, Robinson N, Fischetto G, Dolle G, Alonso JM, Saugy M. Prevalence of Blood Doping in Samples Collected from Elite Track and Field Athletes. *Clin. Chem.* 2011;57:762–9.
- 9.

Doping in elite sports assessed by randomized-response surveys

9. Sparling PB. The Lance Armstrong saga: a wake-up call for drug reform in sports. *Curr. Sports Med. Rep.* 2013;12:53–4.
10. Callaway E. Sports doping: Racing just to keep up. *Nature.* 2011;475:283–5.
11. Pielke Jr R. Gather data to reveal true extent of doping in sport. *Nature.* 2015;517:529.
12. Berry DA. The science of doping: The processes used to charge athletes with cheating are often based on flawed statistics and flawed logic. 2008;454:692–3.
13. Sottas P-E, Saudan C, Saugy M. Doping: a paradigm shift has taken place in testing. *Nature.* 2008;455:166.
14. Baird G. Doping: probability that testing doesn't tell us anything new. *Nature.* 2008;454:692–3.
15. Ljungqvist A, Horta L, Wadler G. Doping: world agency sets standards to promote fair play. *Nature.* 2008;455:1176.
16. Frenger M, Emrich E, Pitsch W. How to produce the belief in clean sports which sells. *Perform. Enhanc. Heal.* [Internet]. Elsevier Ltd; 2013;2:210–5. Available from: <http://dx.doi.org/10.1016/j.peh.2014.09.001>
17. Martensen CK, Møller V. Drugs : Education , Prevention and Policy More money – better anti-doping ? *Drugs Educ. Prev. policy* [Internet]. Informa UK Limited, trading as Taylor & Francis Group; 2016;0:000. Available from: <http://dx.doi.org/10.1080/09687637.2016.1266300>
18. Pitsch W. “The science of doping” revisited: Fallacies of the current anti-doping regime. *Eur. J. Sport Sci.* [Internet]. 2009 [cited 2014 Jun 5];9:87–95. Available from: <http://www.tandfonline.com/doi/abs/10.1080/17461390802702309>
19. Fox JF, Tracy PE. *Randomized response: A method for sensitive surveys.* Lewis-Beck MS, editor. *Quant. Appl. Soc. Sci.* Newbury Park: Sage Publications, Inc.; 1986.
20. Horvitz DG, Greenberg BG, Abernathy JR. Randomized response: A data-gathering device for sensitive questions. *Int. Stat. Rev.* 1976;44:181–96.
21. Lensvelt-Mulders GJLM. Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. *Sociol. Methods Res.* 2005;33:319–48.

22. Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* 1965;60:63–6.
23. Greenberg BG, Abul-Ela A-LA, Simmons WR, G.Horvitz D. The Unrelated Question Randomized Response Model: Theoretical Framework. *J. Am. Stat. Assoc.* 1969;64:520–39.
24. Vital N, Reports S. *National Vital Statistics Reports.* 1999;
25. Ulrich R, Schröter H, Striegel H, Simon P. Asking sensitive questions: A statistical power analysis of randomized response models. *Psychol. Methods [Internet].* 2012 [cited 2014 Jun 4];17:623–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22924599>
26. Chaudhuri A, Christofides TC. *Indirect questioning in sample surveys.* Heidelberg: Springer; 2013.
27. Plessner H, Musch J. Wie verbreitet ist Doping im Leistungssport? Eine www-Umfrage mit Hilfe der Randomized-Response-Technik. In: Strauß B, editor. *Expert. im Sport.* Cologne: bps; 2002. p. 78–9.
28. Pitsch W, Emrich E, Klein M. Doping in elite sports in Germany: results of a www survey. *Eur. J. Sport Soc.* 2007;4:89–102.
29. Striegel H, Ulrich R, Simon P. Randomized response estimates for doping and illicit drug use in elite athletes. *Drug Alcohol Depend.* 2010;106:230–2.
30. Pitsch W, Emrich E. The frequency of doping in elite sport: Results of a replication study. *Int. Rev. Sociol. Sport.* 2012;47:559–80.
31. Dietz P, Ulrich R, Dalaker R, Striegel H, Franke AG, Lieb K, et al. Associations between physical and cognitive doping--a cross-sectional study in 2.997 triathletes. *PLoS One [Internet].* 2013;8:e78702. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3827233&tool=pmcentrez&rendertype=abstract>
32. Schröter H, Studzinski B, Dietz P, Ulrich R, Striegel H, Simon P. A Comparison of the Cheater Detection and the Unrelated Question Models: A Randomized Response Survey on Physical and Cognitive Doping in Recreational Triathletes. *PLoS One [Internet].* 2016;11:e0155765. Available from: <http://dx.plos.org/10.1371/journal.pone.0155765>

Doping in elite sports assessed by randomized-response surveys

33. Frenger M, Pitsch W, Emrich E. Sport-induced substance use-an empirical study to the extent within a German Sports Association. *PLoS One*. 2016;11:1–17.
34. Luchins AS, Luchins EH. *Rigidity of behavior: A variational approach to the effect of Einstellung*. Eugene, OR: University of Oregon Press.; 1959.
35. Tourangeau R, Yan T. Sensitive questions in surveys. *Psychol. Bull.* 2007;133:859–83.
36. Knapik JJ, Steelman RA, Hoedebecke SS, Austin KG, Farina EK, Lieberman HR. Prevalence of Dietary Supplement Use by Athletes: Systematic Review and Meta-Analysis. *Sport. Med.* Springer International Publishing; 2016;46:103–23.

Doping in elite sports assessed by randomized-response surveys

Electronic Supplementary Information is linked to the online version of the paper at [link to be supplied in published text].

Acknowledgements We are grateful to Heba Jamal Al Ardah, Juliana Chialvo Arriola, Derek Jon Galotera, Ricky James, Fatma Lanouar, Dmitri Vorobyov, Ong Chin Wei, and Yoon Ji Yu for their help in data collection. We are especially grateful to Olivier Rabin for launching this project and to Georg Sandberger for legal advice. The authors are also grateful to the International Association of Athletics Federations (IAAF) and the Union of Arab National Olympic Authorities for their permissions to distribute the survey at 13th World Championship in Athletics (Daegu, South Korea, 2011) and at PAG 12th Quadrennial Pan-Arab Games (Doha, Qatar, 2011), respectively. The authors also appreciate the comments of the IAAF's reviewers on previous drafts of the manuscript.

Compliance with Ethical Standards

Funding This work was funded by the World Anti-Doping Agency / Agence Mondiale Antidopage, Montréal, Canada (WADA). The approximate cost of conducting the study was \$155,000 US.

Conflict of interest All authors received travel reimbursement and an honorarium of \$300 Canadian from the World Anti-Doping Agency for their work on designing the study. Andrea Petróczi, Perikles Simon, and Rolf Ulrich have received grant support from WADA for work on other studies. Léa Cléret was employed by WADA when this survey was conducted. Harrison Pope, Tamás Nepusz, Jay Schaffer, Gen Kanayama, and R. Dawn Comstock declare no competing financial interests.

We would note that there was a delay of nearly 6 years between the completion of the data collection and the publication of this paper, due to negotiations between WADA and the IAAF, and subsequently between the IAAF and the authors, regarding the authority to publish

Doping in elite sports assessed by randomized-response surveys

the results. This process has been described in reports in the popular media. However, the authors declare that neither WADA nor the IAAF has imposed any restrictions on the authors' access to the data, the analysis and interpretation of the data, or the writing of this report.

Ethical approval This study was conducted under the approval of the Kingston University Science Faculty Ethics Committee (Kingston on Thames, UK), and has been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Figure legends:

Figure 1. Probability tree for the unrelated question model (UQM).

a. The respondent is asked an initial personal question for which only the respondent knows the answer, but for which the probability p of a “no” response and the complementary probability $q = 1 - p$ of a “yes” response in the overall population are known (e.g., a birthdate; see Figure 2). Respondents answering “yes” to the initial question are directed to a subsequent non-sensitive question A, for which the probability π_N of a “yes” answer is again known (e.g., another birthdate). Respondents answering “no” on the initial question are directed to a sensitive question B (e.g., past-year doping), where the probability π_S of a “yes” answer is unknown and represents the target of the investigation. **b.** The expected proportion λ of “yes” answers from the total group of respondents is therefore a function of p , π_N , and π_S . In the present study, the probability p of being directed to the sensitive question was approximately 2/3. In addition, the probability π_N of answering the non-sensitive question with “yes” was approximately 1/2. Using these values and solving for π_S , the prevalence of past-year doping would be expected to be approximately $(\lambda - 0.167)/0.667$. See the main text for exact numbers.

Figure 2: Sequential screens of the tablet-computer survey instrument.

Athletes received an initial screen (not shown; used at PAG only) in which they could select their sport or alternatively decline to participate in the study. On the next screen (also not shown) athletes could select their preferred language. Thereafter, they were shown a series of screens (shown in the Figure) where they were **a.** asked to consent for participation in the study; **b.** asked to choose a random individual whose birthday was known to them; and **c.** instructed to answer one of two possible questions based on the birthday of the chosen individual. The computer then presented: **d.** a screen showing the two questions, where

Doping in elite sports assessed by randomized-response surveys

question “A” concerned a non-sensitive issue for which the expected proportion of “yes” answers was known, and where question “B” concerned the sensitive issue under study (past-year doping). This screen contained a phrase reminding respondents that no one could know which of the two questions they were answering. At the PAG event, respondents also received a second series of screens identical to screens **b-d**, but with a question about use of supplements substituted for the doping question (screens not shown; see text).

Figure 3. Estimated prevalence of past-year doping with various proportions of fast responders removed. Estimated past-year prevalence (SE) of doping (red, green) and use of supplements (black) at WCA and PAG as a function of deleting 0%, 10%, 20%, 30%, 40%, or 50% of the fastest responders from the total sample.

Figure 1

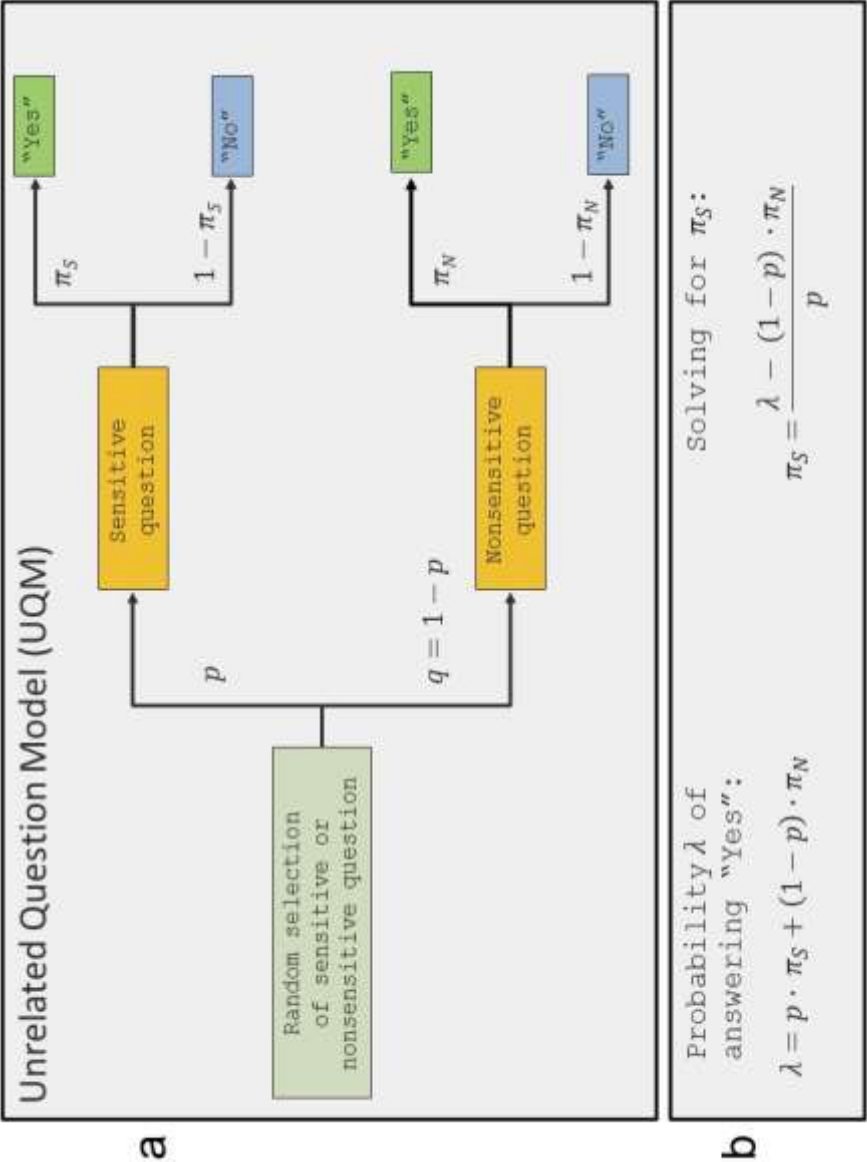


Figure 2

a

Thank you for taking part in this three-step survey. It will take less than 2 minutes of your time.

It is very important to answer truthfully.

No one will ever know what you answered apart from you.

Are you ready to start?

b

Think of someone close to you (it can be anyone, such as your parent, sibling, partner, friend or even yourself) whose date of birth you know.

Are you ready to proceed?

c

Now think about the date of birth of the person you have chosen.

If the date is between the 1st and 10th day of a month, proceed to Question A and please answer it honestly.

If the date is between the 11th and 31st day of a month, proceed to Question B and please answer it honestly.

d

Now think about the date of birth of the person you have chosen.

If the date is between the 1st and 10th day of a month, proceed to Question A and please answer it honestly.

If the date is between the 11th and 31st day of a month, proceed to Question B and please answer it honestly.

Question A: Is the person's date of birth in the first half of the year (January through June inclusive)?

Question B: Have you knowingly violated anti-doping regulations by using a prohibited substance or method in the past 12 months?

Note that only you can know which of the two questions you are answering!

Figure 3

