

Reliability of Cycling Gross Efficiency Using the Douglas Bag Method

JAMES G. HOPKER¹, SIMON A. JOBSON¹, HANNAH C. GREGSON¹, DAMIAN COLEMAN², and LOUIS PASSFIELD¹

¹Centre for Sports Studies, University of Kent, Kent, UNITED KINGDOM; and ²Department of Sports Science, Leisure and Tourism, Canterbury Christ Church University, Kent, UNITED KINGDOM

ABSTRACT

HOPKER, J. G., S. A. JOBSON, H. C. GREGSON, D. COLEMAN, and L. PASSFIELD. Reliability of Cycling Gross Efficiency Using the Douglas Bag Method. *Med. Sci. Sports Exerc.*, Vol. 44, No. 2, pp. 290–296, 2012. **Purpose:** The aim of this study was to establish the reliability of gross efficiency (GE) measurement (the ratio of mechanical power input to metabolic power output, expressed as a percentage) using the Douglas bag method. **Methods:** The experiment was conducted in two parts. Part 1 examined the potential for errors in the Douglas bag method arising from gas concentration analysis, bag residual volume, and bag leakage or gas diffusion rates. Part 2 of this study examined the within-subject day-to-day variability of GE in 10 trained male cyclists using the Douglas bag method. Participants completed three measurements of GE on separate days at work rates of 150, 180, 210, 240, 270, and 300 W. **Results:** The results demonstrate that the reliability of gas sampling is high with a coefficient of variation (CV) <0.5% for both O₂ and CO₂. The bag residual volume CV was ~15%, which amounts to +0.4 L. This could cause the largest error, but this can be minimized by collecting large gas sample volumes. For part 2, a mean CV of 1.5% with limits of agreement of +0.6% in GE units, around a mean GE of 20.0%, was found. **Conclusions:** The Douglas bag method of measuring expired gases and GE was found to have very high reliability and could be considered the gold-standard approach for evaluating changes in GE. Collecting larger expired gas samples minimizes potential sources of error. **Key Words:** POWER OUTPUT, OXYGEN CONSUMPTION, CADENCE, SUBMAXIMAL EXERCISE, MEASUREMENT ERROR

Changes in gross efficiency (GE), defined as the ratio of mechanical power output to metabolic power input and expressed as a percentage, have been shown to correlate with changes in cycling performance (14). Subsequently, Moseley and Jeukendrup (12) calculated that a 1% improvement in GE would equate to a 63-s improvement in 40-km time trial time. This calculation is consistent with the findings of Hettinga et al. (5), who have shown that a change in GE of 0.9% results in a 25.6-s change in 20-km time trial time. Because two studies report that GE has a coefficient of variation (CV) of approximately 4.3%, the ability to reliably measure such changes remains unclear (12,13). However, both studies used online breath-by-breath gas analysis systems, whereas Carter and Jeukendrup (3) consider the Douglas bag method to be the gold-standard approach. Indeed, in 1955, Taylor et al. (16) reported the error associated with Douglas bag testing of $\dot{V}O_{2\max}$ to be 2.4%. The complexity of online breath-by-breath systems

means researchers rely on a simple initial calibration procedure to ensure the accuracy and reliability of their measurements. In contrast, the much simpler Douglas bag method relies more heavily on first-principle procedures, each of which can be separately evaluated. Therefore, the major sources of error with the Douglas bag method can be identified and quantified. To our knowledge, the reliability of the Douglas bag method and its use in the assessment of submaximal expired gases and GE have not been reported.

The calculation of GE relies on the accurate measurement of power output and power input. The accuracy and reliability of power output systems have been variously reported and are known to have only small implications for the measurement of GE (e.g., Jones and Passfield [11]). In contrast, the reliability of expired gas measurement to estimate power input and consequent errors in calculation of GE seem much worse. Moseley and Jeukendrup (12) and Noordhof et al. (13) demonstrated a mean CV of 4.2%–4.4% for GE and suggested that they can detect changes of ~0.6% in GE using an online gas analysis system. Given the large effects of a small change in GE on performance, the highest possible precision of measurement is desirable. Similarly, to evaluate the effects of selected interventions on performance, repeatability of key measurements must be high, and within-participant variation must be low. Thus, examining the reliability of measuring GE using the Douglas bag method will permit appropriate sample and effect sizes for subsequent experiments to be established (1,9). Therefore,

Address for correspondence: Louis Passfield, Ph.D., Centre for Sports Studies, University of Kent, Kent ME4 4AG, United Kingdom; E-mail: L.Passfield@kent.ac.uk.

Submitted for publication April 2011.

Accepted for publication June 2011.

0195-9131/12/4402-0290/0

MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Copyright © 2012 by the American College of Sports Medicine

DOI: 10.1249/MSS.0b013e31822cb0d2

the purpose of this study was to determine the reliability of the Douglas bag method of gas analysis and its use in the measurement of GE in cycling. The study was divided into two parts. Experiment 1 was concerned with the assessment of the reliability of the Douglas bag method (open-circuit spirometry). In particular, the aim was to quantify the effect of variables thought likely to influence the accurate measurement of expired gas concentration and volume. Experiment 2 was used to assess the reliability of GE using the Douglas bag method.

METHODS

Experiment 1: Calibration of the Douglas Bag Open-Circuit Spirometry Method

This experiment examined the reliability of the gas sampling procedure to determine its inherent variability. In addition, the influence of the residual volume of a Douglas bag after evacuation on the subsequently collected gas concentration was determined. Also, the rate of leakage or diffusion of collected gases from the Douglas bag was measured. All experimental work was conducted after university ethical approval and after obtaining informed consent from participants.

Sampling Reliability

The expirate from an individual performing moderate-intensity exercise was collected via a Hans Rudolph breathing valve (2700; Hans Rudolph, Inc., Kansas City, MO) and plastic tubing into a plastic Douglas bag (Plysu Industrial, Ltd., Milton Keynes, UK) for repeated gas analysis. The concentration of O₂ and CO₂ were repeatedly determined from this bag on 20 separate occasions to determine the variability in sampling. During repeated sampling, the gas analyzers were running continuously and were recalibrated after analysis of 10 samples.

Residual Volume

The residual volumes of 13 different Douglas bags and one Douglas bag on six separate occasions were determined. The residual volume was determined by gas dilution. This method was preferred to volumetric measurement because variability in residual volume was thought to affect measured gas concentrations more profoundly than gas volume. Each Douglas bag was used to collect approximately 50 L of expirate from a participant undertaking moderate-intensity exercise. The Douglas bags were subsequently analyzed for O₂ and CO₂ concentrations and evacuated with a vacuum pump following normal laboratory procedures. Immediately after evacuation, a Hans Rudolph gas syringe (Hans Rudolph, Inc.) was used to introduce 7 L of outside air into the Douglas bag. The gas concentrations in the Douglas bag were then reanalyzed. The residual volume was determined measuring the changes in O₂ and CO₂ concentrations. The

7-L air sample was gathered from outside the building, away from any possible contaminating ventilation exhaust systems, and was assumed to consist of 20.93% O₂ and 0.03% CO₂. Particular care was also taken with the 7-L syringe procedure that was connected to the Douglas bag via a two-way Salford respiratory valve box (Cranlea, Birmingham, UK). The airtight operation of this valve and all connections between it and the Douglas bag were verified before use. As an additional precaution, the valve not in use when syringing (i.e., the inlet or outlet valve as appropriate) was sealed to ensure that only the intended gas volume could pass through the system.

Gas Exchange between Douglas Bag and Ambient Air

The rate of exchange of O₂ and CO₂ between a Douglas bag and the laboratory environment was measured by periodically determining the gas concentration of the bag during a period of 147 h. Approximately 70 L of expirate from a participant engaged in moderate-intensity exercise was collected in a Douglas bag. The participant adopted a slow, deep breathing pattern to maximize the respective changes in fraction of expired O₂ and fraction of expired CO₂ (F_EO₂ and F_ECO₂), respectively.

The next part of the experiment examined the agreement between two different methods of determining gas volume. A 7-L gas syringe (Hans Rudolph, Inc.) was used to produce and compare known gas volumes with a dry gas volume meter (Harvard Apparatus, Ltd., Edenbridge, UK) used during standard expired gas volume measurements. The syringe method was found to be highly reproducible when used to fill and then empty a Douglas bag. The repeated measurements obtained with the 7-L syringe agreed to within 50 mL, irrespective of the volume, over a range of 10 to 150 L.

The 7-L syringe was used to carefully introduce a range of known volumes (10–160 L) of ambient air into a Douglas bag. The system and procedure adopted for syringing were as described above for determining residual volume. The Douglas bag was immediately evacuated through the dry gas volume meter with a vacuum pump at a flow rate of 60 L·min⁻¹. The system was sealed (by blocking the air outlet) while the vacuum pump was started to check no leaks existed and to help maintain a constant rate of flow during Douglas bag evacuation. Once empty, the Douglas bag was gently manipulated to help expel as much air as possible. A residual volume was consistently found after evacuation by further emptying with the syringe. Accordingly, two trials were conducted: first, with the residual volume determined for each metered volume, and second, to replicate normal laboratory practice agreement, trials were undertaken with the residual volume ignored.

Experiment 2: Reliability of the Measurement of GE

This study examined the variability in the measurement of GE during cycling using the Douglas bag method. After

university ethical approval, 10 male cyclists (mean \pm SD: age = 36 \pm 9 yr, mass = 77 \pm 9 kg, maximal aerobic power (MAP) = 366 \pm 30 W, $\dot{V}O_{2peak}$ = 59.0 \pm 8 mL \cdot kg $^{-1}$ \cdot min $^{-1}$) with at least 2 yr of training history provided written informed consent to participate. All participants were asked to follow their normal diet throughout the duration of the study and requested not to train or consume any alcohol or caffeine less than 24 h before testing.

Participants visited the laboratory on four separate occasions, at the same time of day and on the same days of the week. During the first visit, a progressive maximal cycle ergometry test was conducted to calculate maximal aerobic power, whereas the following three visits were used to repeatedly measure GE. All tests were conducted on an SRM cycle ergometer (Schoberer Rad Messtechnik, Jülich, Germany). Participants' weight and height were measured using a seca beam balance scale and stadiometer, respectively (seca, Hamburg, Germany).

Maximal Aerobic Power Test

Before testing, participants performed a 10-min warm-up at 100 W. After the warm-up, the required power output was increased by 20 W every minute. The test continued until volitional exhaustion when participants were unable to produce the required power output. Participants' $\dot{V}O_{2peak}$ was established via the collection of expired gases in the last minute of the test using Douglas bags (Hans Rudolph, Inc.). When a participant indicated he or she had approximately 1 min of exercise remaining, gas collection was started with a stopwatch timing the duration the bag was open. The $F_{E}O_2$ of the expired gas was subsequently analyzed using a high-accuracy gas analyzer (Servomex, West Sussex, UK). The volume of expired air collected in the Douglas bags was analyzed using a dry gas meter (Harvard Apparatus, Kent, UK). Both the $F_{E}O_2$ and the bag volume were then used to calculate the $\dot{V}O_2$ on the basis of the sampling time. Maximal aerobic power output was established as the average power output recorded by the ergometer during the last minute of the test. Participants were asked to maintain a constant but individually chosen cadence throughout the test.

GE Tests

During the second, third, and fourth visits, participants performed repeated tests of cycling GE. The method of Passfield and Doust (14) was used to measure GE, which was calculated as the ratio of power output to power input. For every test, the participants' normal bicycle riding position was replicated on the ergometer. Participants initially completed a 10-min warm-up at 100 W using their preferred cadence, which was established during the maximal aerobic power test (described above). After the warm-up, participants' GE was measured at work rates of 150, 180, 210, 240, 270, and 300 W. The different work rate stages were randomly ordered and lasted 6 min, with a 5-min rest

between stages. Expired air was collected by the Douglas bag method between the fifth and sixth minutes of each work stage. Subsequently, the $F_{E}O_2$ and $F_{E}CO_2$ of expired air were measured using a high-accuracy gas analyzer (Servomex). The expired volume of air was measured using a dry gas meter (Harvard Apparatus, Kent, UK). Power input was calculated from the $\dot{V}O_2$ and its energetic equivalent according to the table of nonprotein respiratory quotient (15). Power output was recorded as the average power output for the last minute of each stage.

Data Analysis

Experiment 1. Before any analysis, all data were assessed for normality of distribution and heteroscedasticity. The variability of repeated gas samples was assessed for random measurement error by the use of the CV. The linearity of the relationship between the two methods for gas volume measurement was determined by scatterplot and correlation coefficient (r). Thereafter, a calibration equation was obtained by linear regression. Finally, the agreement between the two methods was examined by determining the limits of agreement (2) from the residuals of the calibration equation.

Experiment 2. GE for each of the three tests at each work rate was calculated. Data in which an RER >1.0 was found were removed before analysis. The individual typical error was expressed as a CV calculated from the GE data for each participant at each work rate across each of the three repeated visits. Mean data across all three trials at each power output was assessed using the root mean square error. Confidence intervals (95% CI) of the CV and 95% limits of agreement were calculated per participant to assess the variability of the repeated tests (9). Comparisons of the GE at the different intensities across days were assessed using repeated-measures ANOVA; statistical significance was set at 95% confidence ($P < 0.05$). All values are expressed as mean \pm SD unless otherwise stated.

RESULTS

Experiment 1: Calibration of Open-Circuit Spirometry System

Sampling reliability. Results for the 20 repeated gas samples demonstrated that the mean measured O_2 was 16.18 and CO_2 was 4.44, with 95% CIs of 16.17–16.20 and 4.40–4.48 for O_2 and CO_2 , respectively. The O_2 analyzer

TABLE 1. Calculated residual volume for 13 different Douglas bags and six repeated measures on one Douglas bag based on dilution of % O_2 and % CO_2 .

	Residual Volume from % O_2 (L)		Residual Volume from % CO_2 (L)	
	13 Bags	1 Bag	13 Bags	1 Bag
Mean	1.487	2.051	1.552	2.045
SD	0.228	0.164	0.226	0.204
CV (%)	15.4	8.0	14.6	9.9

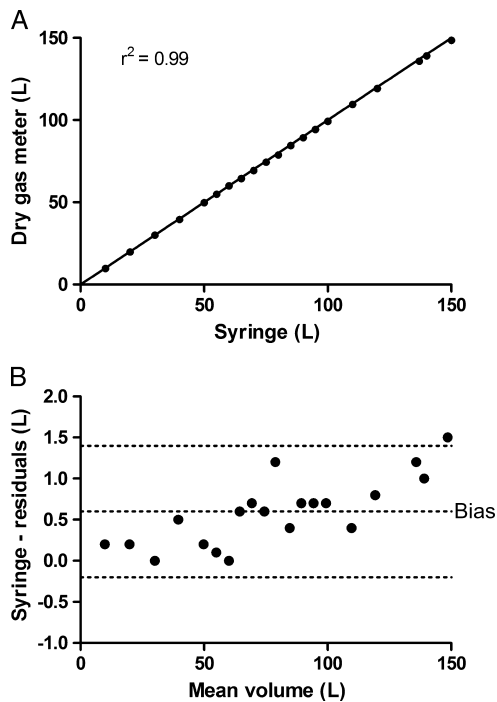


FIGURE 1—Gas volume reliability measurements. *Top panel:* the high linear correlation between syringe and gas volume meter values ($r^2 = 0.99$) where residual volume was ignored. • indicates gas meter reading; — indicates line of identity. *Bottom panel:* Bland–Altman plot of difference between syringe and corrected meter volumes, with 95% limits of agreement shown.

exhibited slightly less variability than the CO₂ analyzer as evidenced by the respective CVs of 0.05% and 0.45%.

Residual volume. The gas concentration of the residual volume is assumed to be the same as that of the bag before evacuation. If no residual volume was present, concentrations of 20.93% and 0.03% for O₂ and CO₂ were expected after adding 7 L of outside air. Hence, any change from these concentrations reflected a dilution caused by the residual volume. The mean residual volume from 13 separate Douglas bags was 1.487 or 1.582 L as determined from changes in %O₂ and %CO₂, respectively. The mean, SD, and CV for all 13 Douglas bags and for the repeated measures on one bag are provided in Table 1.

Gas leakage or diffusion from Douglas bag. An essentially linear relationship between time and corresponding changes in Douglas bag gas concentrations for both O₂ and CO₂ was observed. The rate of O₂ loss from the Douglas bag was slower than CO₂ with concentration changes of 0.005%·h⁻¹ and -0.015%·h⁻¹, respectively.

Gas volume measurement. Agreement between the 7-L gas syringe and dry gas meter was high, with 95% limits of agreement of ±0.82 L (SD = 0.42 L) for the raw volume data and ±0.49 L (SD = 0.25 L) for the residuals from the calibration equation. The raw score differences had a bias of 0.59 L, which was removed by linear regression (Fig. 1, top panel). A Bland–Altman plot is shown for the residuals from the calibration equation (Fig. 1, bottom panel), which indicates a small positive bias as the volume increases.

Experiment 2

Mean GE results for the group of participants were 18.1% ± 1.2%, 18.8% ± 1.0%, 19.3% ± 1.2%, 20.0% ± 1.0%, 20.1% ± 0.8%, and 19.9% ± 0.8% for work rates of 150, 180, 210, 240, 270, and 300 W, respectively.

A mean group CV of 1.3% (95% CL = 0.9%–2.5%) for trials 2–1 and 1.2% (95% CL = 0.9%–2.3%) for trials 3–2 across all six power outputs was found. Table 2 provides a summary of the trial-to-trial data across all common work rates that participants successfully completed on all three visits. There was no significant difference in the GE measures across trials ($P > 0.05$), and trial 3–2 comparisons all fell within the 95% CI of the trial 2–1 comparisons. On the basis of a GE of 20%, the limits of agreement were ±0.7% in GE units or ±3.6% of the measure for trials 2–1 and ±0.7% to ±3.3% for trials 3–2. Because there was no evidence of change in GE across the three trials, a single %CV was derived using the root mean square error from ANOVA. This process derived a typical error expressed as a CV (%) of 1.5% (95% CL = 1.1%–2.2%), with the 95% limits of agreement being ±2.9% or ±0.6 GE units.

Figure 2 illustrates the agreement between the three repeated trials across all work rates in this group of cyclists.

TABLE 2. Mean intertrial results for the cyclist group across all absolute work rates at preferred cadence, as well as mean across all common work rates for each individual for trials 2–1 and 3–2.

	150 W		180 W		210 W		240 W		270 W		300 W		Mean across All Common Work Rates	
	Trial 2–1	Trial 3–2	Trial 2–1	Trial 3–2	Trial 2–1	Trial 3–2	Trial 2–1	Trial 3–2	Trial 2–1	Trial 3–2	Trial 2–1	Trial 3–2	Trial 2–1	Trial 3–2
Typical error as a CV (%)	2.2	2.0	1.9	2.7	1.8	1.3	2.6	1.4	1.8	1.5	2.6	1.7	1.3	1.2
Lower 95% CL	1.5	1.4	1.3	1.9	1.3	0.9	1.7	0.9	1.0	0.8	1.3	0.8	0.9	0.9
Upper 95% CL	4.1	3.8	3.5	5.1	3.4	2.3	6.6	3.6	6.8	10.0	17.5	71.2	2.5	2.3
LOA (%) based on %CV	6.1	5.5	5.3	7.5	5.0	3.6	7.2	3.9	4.5	4.2	7.2	4.7	3.6	3.3
LOA (%) based on GE of 20%	1.2	1.1	1.1	1.5	1.0	0.7	1.4	0.8	1.0	0.8	1.4	0.9	0.7	0.7

LOA, limits of agreement.

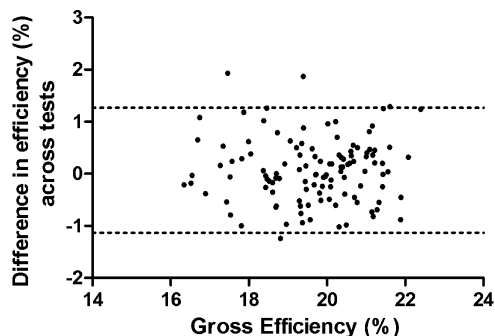


FIGURE 2—Agreement between repeated tests of GE in trained cyclists using their preferred cadence.

Repeated-measures ANOVA demonstrated no significant change over time ($P > 0.05$).

DISCUSSION

Experiment 1. It is evident from the results of this study that the repeatability of the gas sampling procedure is extremely high. A slightly greater variability was experienced for CO_2 than O_2 , but in both situations, the CV was less than 0.5%. Factors that may influence the CV are the stability of the gas analyzers, the reliability of the recalibration procedure, and possible variation in flow rate through the gas analyzers during sampling. The very low CVs suggest that these variables will not lead to a significant error during repeated gas measurements under normal laboratory conditions.

The residual volume of all the Douglas bags in the laboratory was determined by gas dilution. After evacuation with a vacuum pump, typically, 1.5 L of air still remained in the Douglas bag. The difference in residual volume as determined by $\% \text{O}_2$ and $\% \text{CO}_2$ is probably due to the different resolution of the respective analyzers and the precision offered by the varying magnitude of change in the residual and ambient gas concentrations. The variability of the residual volume was rather large ($\text{CV} \approx 15\%$); however, in absolute terms, this amounts to the measured volume of 95% of Douglas bags varying by ± 0.4 L. This magnitude of difference is unlikely to produce meaningful errors in the calculation of either $\dot{V}\text{O}_2$ or \dot{V}_E (± 0.02 and ± 0.4 L, respectively, with a 30-L sample) and thus GE. Because the residual volume approximates a fixed error, its effects are influenced by the size of the sample volume collected. Therefore, it is recommended that researchers always collect the largest gas sample volume possible (e.g., by extending duration of gas collection) and exercise extreme care where the collection of a small sample volume is unavoidable.

The major implication of the variable Douglas bag residual volume is related to its contamination of a subsequent gas sample. It is for this reason that residual volumes were determined by gas dilution rather than by gas syringe in the present study. For example, a residual volume composed entirely of ambient air ($\text{O}_2 = 20.93\%$) will tend to increase the apparent $\% \text{F}_E \text{O}_2$ of a subsequent collected expired gas sample. The theoretical consequences of this dilution effect are explored in Table 3. Example 1 demonstrates that a 40-L Douglas bag expired gas sample with an $\text{F}_E \text{O}_2$ of 14.5% mixed with a typical residual volume of 18.0% O_2 concentration would result in a (measured) Douglas bag $\% \text{O}_2$ concentration of 14.63%, causing a difference of 0.13%. This difference translates into an error in calculated $\dot{V}\text{O}_2$ of approximately $0.1 \text{ L} \cdot \text{min}^{-1}$. Consequently, the effect of residual volume gas contamination can be seen to be more than twice that of the error in measured volume alone. Further, it is important to note that simply increasing the expired gas sample volume collected in the Douglas bag will markedly reduce this contamination error. Examples 2 and 3 in Table 3 illustrate the effect of changes of ± 2 SD in residual volume on an $\text{F}_E \text{O}_2$ value (16.25%). Example 4 demonstrates that the greatest error and implication for GE measurement are created if the residual volume is composed of ambient air. When considering the effect of Douglas bag residual volume contamination error in the measurement of GE, examples 1, 2, and 4 result in changes of >1 SD of the repeated GE trials found in experiment 2. James and Doust (10) reported that the CV of $\dot{V}\text{O}_2$ determination during moderate-intensity treadmill running is only 1.4%. However, the effect of the Douglas bag residual volumes found in this experiment could potentially account for a large proportion of this CV. However, the residual volume contamination can be minimized by “flushing” Douglas bags with expirate before use and collecting large expired gas sample volumes.

The rates of gas leakage or diffusion from the Douglas bags for both O_2 and CO_2 were both found to be slow, although CO_2 was more rapid. The rate of loss of both gases per hour was below the resolution of their respective gas analyzers. Therefore, periods of <1 h between gas sample collection and analysis do not seem to result in meaningful changes in the measured gas concentrations.

This study has found excellent agreement between expired gas volume measures determined with a 7-L gas syringe and a dry gas volume meter, with 95% of the differences falling within ± 0.49 L. Even gas sample volumes measured without establishing a calibration slope for the dry gas meter provide an acceptable level of agreement of less than ± 1 L.

TABLE 3. Effect of changes in residual volume and its gas concentration on measured concentration of $\% \text{O}_2$ for a 40-L Douglas bag sample.

Example No.	Actual $\% \text{O}_2$	Residual Volume	Residual $\% \text{O}_2$	Measured $\% \text{O}_2$	Absolute $\% \text{O}_2$ Difference	Difference in %GE
1	14.50	1.487	18.00	14.63	0.13	0.40%
2	16.25	1.030	14.50	16.29	0.04	0.26%
3	16.25	1.944	14.50	16.33	0.08	0.54%
4	16.25	1.487	20.93	16.42	0.17	1.17%

Efficiency calculation assumes unchanged work rate and $\dot{V}\text{CO}_2$.

TABLE 4. Effect of a Douglas bag method of expired gas collection on required sample size.

Study	Design	Sample Size	Finding (Values Shown as an Absolute Change in GE)	Required Douglas Bag Sample	Effect Size
Hopker et al. (6)	Longitudinal	14 trained cyclists	GE increases significantly by 1.34% during the pre-season period of a competitive cycling season	4	3.03
Hopker et al. (7)	Cross-sectional Independent groups	32 in two groups (16 trained and 16 untrained)	GE significantly higher in trained cyclists at: 150 W = 1.02% 50% $\dot{V}_{O_{2max}}$ = 1.66% 60% $\dot{V}_{O_{2max}}$ = 1.70%	10 7 5	1.35 1.69 2.15
Hopker et al. (8)	Within groups	29 trained cyclists in two groups (14 high-intensity training group, 15 control group)	GE significantly increases by 1.50% after high-intensity training	5	1.84

Sample size calculations are based on the methods outlined by Baguley; 2004 (1a) and Cohen; 1988 (3a) using GPower software (Erdfelder et al.; 1996 (4a)). Required Douglas bag sample size represents the number of participants required per group to demonstrate the change in GE found in the cited study. Effect size was calculated as the typical error from the highest common work rate across all repeated tests using the Douglas bag method in the current study, divided by the pooled SD observed in the studies cited in the table above (6–8).

Furthermore, these limits of agreement also include the influence of error within the calibration syringe and any changes in the residual volume of the Douglas bag as discussed above. Consequently, the error in gas volume measurement is unlikely to be meaningful provided a reasonable Douglas bag sample gas volume is collected (e.g., >30 L).

Experiment 2. The total within-subject variation in GE found in this study was 1.5%. This value is notably less than previously reported for online breath-by-breath gas analysis systems (12,13). Moseley and Jeukendrup (12) used a graded exercise protocol with increments every 3 min and found a mean within-subject CV of 4.2%. Noordhof et al. (13) obtained similar CV values (4.4%) from 6-min stages at 45%, 55%, and 65% of participants' power output at $\dot{V}_{O_{2max}}$. The mean CVs calculated in the present study suggest an improvement in GE units as small as 0.4% can be reliably detected in trained cyclists. Table 4 uses data from the current study to recalculate the number of participants required to detect significant differences in GE reported in previously published studies from our laboratory (6–8). These previous studies have found that there are significant differences in GE between trained and untrained cyclists (7) and also that GE changes during a competitive season (6) and in response to high-intensity training (8). Typically, these studies required approximately 30 trained cyclists to detect these changes in GE using an online breath-by-breath gas analysis system. The reliable Douglas bag technique evaluated in the present study may make it possible to detect a significant change/difference in GE with considerably lower participant numbers than these previous studies.

The findings of the present study agree with Carter and Jeukendrup's (3) suggestion that the Douglas bag method should be considered the "gold standard." Because the

Douglas bag method uses a largely first-principle-based approach, it minimizes the assumptions required compared with online breath-by-breath systems. Online systems process expired gases in real time, and thus, errors may occur in the measurement of volume or concentration with every breath. Indeed, nonlinear responses, in particular at low and high flow rates, have been reported with online systems (17). With the Douglas bag method, the scope for such errors may be limited. Furthermore, many online systems assume that expired air will be saturated and have a temperature of ~32°C. Therefore, unlike the Douglas bag method, these are unable to account for differences in temperature and water vapor pressure, which may also lead to errors in the calculated gas concentrations (4).

CONCLUSIONS

In conclusion, this study has demonstrated that the procedures for determining the concentrations of expired air samples using the Douglas bag method show high reliability. However, a persistent residual volume in Douglas bags has been found and may create a notable error by contaminating a subsequent gas sample. This error may be minimized by "flushing" the Douglas bag and working with large expired gas sample volumes. The high reliability of the Douglas bag method resulted in low within-subject variability in the measurement of GE. A change in GE of as small as 0.4% may therefore be reliably detected. Consequently, it is recommended that the Douglas bag method be used to evaluate differences or changes in GE, particularly where these are small.

No external funding was sought or received for this work.

The authors report no conflicts of interest.

The results of the present study do not constitute endorsement by the American College of Sports Medicine.

REFERENCES

- Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217–38.
- Baguley T. Understanding statistical power in the context of applied research. *Appl Ergon.* 2004;35:73–80.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307–10.
- Carter J, Jeukendrup AE. Validity and reliability of three commercially available breath-by-breath respiratory systems. *Eur J Appl Physiol.* 2002;86(5):435–41.
- Cohen J. *Statistical Power Analysis for the Behavioural Sciences* (2nd Ed.). London: Lawrence Erlbaum Associate Publishers, 1988.
- Elia M, McDonald T, Crisp A. Errors in measurements of CO₂ through the use of drying agents. *Clin Chim Acta.* 1986;158(3):237–44.

- 4a. Erdfelder E, Faul F, Buchner A. GPower, a general power analysis programme. *Behav Res Methods*. 1996;28:1–11.
5. Hettinga FJ, De Koning JJ, de Vrijer A, Wüst RC, Daanen HA, Foster C. The effect of ambient temperature on gross-efficiency in cycling. *Eur J Appl Physiol*. 2007;101(4):465–71.
6. Hopker JG, Coleman DA, Passfield L. Changes in cycling efficiency during a competitive season. *Med Sci Sports Exerc*. 2009; 41(4):912–9.
7. Hopker JG, Coleman DA, Wiles JD. Differences in efficiency between trained and recreational cyclists. *Appl Physiol Nutr Metab*. 2007;32(6):1036–42.
8. Hopker JG, Coleman DA, Passfield L, Wiles JD. The effect of training volume and intensity on competitive cyclists' efficiency. *Appl Physiol Nutr Metab*. 2010;35(1):17–22.
9. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30(1):1–15.
10. James DV, Doust JH. Oxygen uptake during moderate intensity running: response following a single bout of interval training. *Eur J Appl Physiol Occup Physiol*. 1998;77(6):551–5.
11. Jones SM, Passfield L. The dynamic calibration of bicycle power measuring cranks. In: Haake SJ, editor. *The Engineering of Sport*. Oxford (UK): Blackwell Science; 1998. p. 265–74.
12. Moseley L, Jeukendrup AE. The reliability of cycling efficiency. *Med Sci Sports Exerc*. 2001;33(4):621–7.
13. Noordhof DA, de Koning JJ, van Erp T, et al. The between and within day variation in gross efficiency. *Eur J Appl Physiol*. 2010;109(6):1209–18.
14. Passfield L, Doust JD. Changes in cycling efficiency and cycling performance after endurance exercise. *Med Sci Sports Exerc*. 2000;32(11):1935–41.
15. Péronnet F, Massicotte D. Table of nonprotein respiratory quotient: an update. *Can J Sport Sci*. 1991;16(1):23–9.
16. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J Appl Physiol*. 1955;8(1):73–80.
17. Yeh MP, Adams TD, Gardner RM, Yanowitz FG. Turbine flowmeter vs. Fleisch pneumotachometer: a comparative study for exercise testing. *J Appl Physiol*. 1987;63(3):1289–95.