Evaluation of the predictive performance of four pharmacokinetic models for propofol

J. B. Glen1* and F. Servin2

1Research Department, Glen Pharma Ltd, 35A Bexton Road, Knutsford, Cheshire, UK. 2Department of Anaesthesia, University Hospital Centre, Hopital Bichat, 75877 Paris, France

*Corresponding author. E-mail: jbg@glenpharma.com

Background. This study has compared the predictive performance of four pharmacokinetic models, two of which are currently incorporated in commercial target-controlled infusion pumps for the administration of propofol.

Methods. Arterial propofol concentrations and patient characteristic data were available from nine patients who, in a published study, had received a standardized infusion of propofol. Predicted concentrations with ‘Diprifusor’ (Marsh), ‘Schneider’, ‘Schuttler’, and ‘White’ models were obtained by computer simulation. The predictive performance of each model was assessed overall and over the following phases: rapid infusion (1–5 min), early (1–21 min), maintenance (21-min end-infusion), and recovery (2–20 min post-infusion).

Results. The overall assessment, based on 29–36 samples from each patient, indicated that all four models were clinically acceptable. However, the negligible bias (<0.1%) with the ‘Schneider’ model was accompanied by overprediction in the rapid infusion phase and underprediction during recovery. This changing bias over time was not detected as ‘divergence’ when assessed on absolute performance error (APE), (1.4% h⁻¹) but became significant (13.2% h⁻¹) when based on changes in signed PE over time. The ‘Schuttler’ model performed well at most phases but overpredicted concentrations during recovery. The White model led to a marginal improvement over ‘Diprifusor’ and would be expected to reduce the positive bias usually seen with ‘Diprifusor’ systems.

Conclusions. In assessing the predictive performance of pharmacokinetic models, additional information can be obtained by analysis of bias at different phases of an infusion. The evaluation of divergence should involve linear regression analysis of both absolute and signed PEs.

Br J Anaesth 2009; 102: 626–32

Keywords: anaesthetics i.v., propofol; computer simulation; drug delivery, computerized; pharmacokinetics, models; pharmacokinetics, propofol

Accepted for publication: February 9, 2009

The first commercial target-controlled infusion (TCI) devices became available in 1996 and all incorporated the ‘Diprifusor’ TCI module (AstraZeneca, Macclesfield, UK), which uses the Marsh model of the pharmacokinetic model described by Gepts and colleagues. A typographical error occurred in the Marsh publication where the value of $k_{12}$ implemented in Diprifusor software is 0.114 min⁻¹ as in the original Gepts paper. In this study all simulations were performed with the ‘Diprifusor’ model instead of the Marsh model which has $k_{12}$ of 0.112 min⁻¹.) This model was selected for clinical studies, on the basis of simulation studies, as the most accurate of the three models (‘Marsh’, ‘Tackley’, and ‘Dyck and Shafer’) evaluated at that time. The same three models were compared in a clinical study and similar results obtained. By the selection of a single preferred model, the delivery of propofol in any TCI device incorporating the Diprifusor module was standardized in pumps manufactured by different companies. Clinical validation studies with prototype Diprifusor systems provided information on target blood propofol settings for inclusion in propofol (‘Diprivan’, AstraZeneca) drug labelling, and assessment of predictive performance in two studies indicated a degree of positive bias, which was considered clinically acceptable. Diprifusor TCI systems are now widely used in most countries of the world but require the use of an
electronically tagged prefilled syringe of propofol. As less-expensive preparations of propofol have become available, a demand arose for TCI devices that did not require the tagged presentation. Two such systems are the ‘Base Primea’ (Fresenius Kabi, Brezins, France) and the ‘Asena PK’ (Cardinal Health, Runcorn, UK). These systems provide the user with a choice of two models for the administration of propofol, the Marsh model or a population model with covariates as described by Schneider and colleagues. As different models may deliver different amounts of propofol, this study was designed to compare the predictive performance of the Diprifusor and Schneider models for propofol. The study was extended to include a recent modification of the Marsh model proposed by White and co-workers, with covariates for age and sex, and another population model described by Schuttler and Ihmsen.

**Methods**

Computer simulation using the program PK-SIM (Specialized Data Systems, Jenkintown, PA, USA) was used to predict blood propofol concentrations with each pharmacokinetic model. The input profile was that used in an earlier study, which compared the pharmacokinetics of propofol administered as an infusion in patients with cirrhosis and in control patients with normal renal and hepatic function. Of the 10 control patients, this study used data for nine for whom complete patient characteristic information was available. In the clinical study patients had been premedicated orally with diazepam and atropine and anaesthesia induced and maintained using a stepwise infusion of propofol 21 mg kg h\(^{-1}\) for 5 min, 12 mg kg h\(^{-1}\) for 10 min, and 6 mg kg h\(^{-1}\) for the rest of the procedure that lasted for a minimum of 2 h. Small incremental doses of fentanyl (50 μg) were given i.v. as required and the patient’s lungs were ventilated to normocapnia with a mixture of 66% nitrous oxide in oxygen. Predicted propofol concentrations obtained by simulation of the propofol infusion scheme used were compared with arterial blood concentrations obtained by simulation of the propofol infusion and at 2, 4, 6, 8, 10, 20 min after the end of infusion. A variable number of additional samples were collected when the duration of infusion exceeded 120 min.

The predictive performance of each pharmacokinetic model was assessed using the methodology proposed by Varvel and colleagues. At each time point when a measured blood concentration was available, the PE was calculated as:

\[
PE(\%) = \frac{C_m - C_p}{C_p} \times 100
\]

where \(C_m\) and \(C_p\) are the measured and predicted blood concentrations. For each patient, median PE (MDPE) as a measure of bias, and median absolute PE (MDAPE) as a measure of inaccuracy were determined. Values were calculated using all the samples for a given patient and also for the following periods: 1–5 min (rapid infusion), 1–21 min (early phase), 25 min to end of infusion (maintenance phase), and 2–20 min after the end of infusion (recovery phase). The variability in PE was characterized by wobble (the median absolute deviation of PE from MDPE). Divergence was calculated in two ways: as the slope of the linear regression of absolute performance error (APE) against time as advocated by Varvel and colleagues and also as the regression of signed PE against time. Median values obtained with the Diprifusor group were compared with values obtained in the other groups with the Wilcoxon signed rank test. Fisher’s Exact test was used to compare the proportions of patients in which bias of 20% or less was seen. Linear regression was also used to examine the relationship between the duration of the maintenance infusion and the overall value of divergence, the overall value of MDPE and the MDPE for the maintenance phase obtained for each patient. A value of \(P<0.05\) was considered significant. Statistical analysis was performed with the Data Analysis module of Excel (Microsoft) and StatsDirect software (StatsDirect Ltd, Altrincham, UK).

**Results**

The physical characteristics of the nine patients studied are given in Table 1. The duration of infusion exceeded 120 min in all patients. A total of 286 arterial propofol concentrations were compared with concentrations predicted by each of the four models evaluated at each measurement point. Each patient contributed 29–36 samples with five samples from the rapid infusion phase, 13–15 samples from the early phase, 9–15 from the maintenance phase, and 6–7 from the recovery period. Figure 1 provides an illustration of the inter-patient variability in measured blood propofol concentrations with the standardized infusion scheme used. Among the times shown, the greatest degree of variation was seen at the 5 min time-point.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sex</th>
<th>Age (yr)</th>
<th>Body weight (kg)</th>
<th>Height (cm)</th>
<th>LBM (kg)</th>
<th>BMI (kg m(^{-2}))</th>
<th>Duration of infusion (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>M</td>
<td>24</td>
<td>60</td>
<td>172</td>
<td>50.4</td>
<td>20.2</td>
<td>142</td>
</tr>
<tr>
<td>D2</td>
<td>M</td>
<td>34</td>
<td>55</td>
<td>168</td>
<td>46.8</td>
<td>19.5</td>
<td>226</td>
</tr>
<tr>
<td>D3</td>
<td>M</td>
<td>55</td>
<td>70</td>
<td>172</td>
<td>55.8</td>
<td>23.6</td>
<td>287</td>
</tr>
<tr>
<td>H5</td>
<td>M</td>
<td>56</td>
<td>85</td>
<td>170</td>
<td>61.5</td>
<td>29.4</td>
<td>180</td>
</tr>
<tr>
<td>H6</td>
<td>F</td>
<td>52</td>
<td>70</td>
<td>172</td>
<td>50.4</td>
<td>23.6</td>
<td>151</td>
</tr>
<tr>
<td>W2</td>
<td>M</td>
<td>33</td>
<td>67</td>
<td>172</td>
<td>54.3</td>
<td>22.6</td>
<td>133</td>
</tr>
<tr>
<td>W3</td>
<td>M</td>
<td>39</td>
<td>96</td>
<td>184</td>
<td>70.8</td>
<td>28.4</td>
<td>162</td>
</tr>
<tr>
<td>W5</td>
<td>F</td>
<td>30</td>
<td>50</td>
<td>150</td>
<td>37.1</td>
<td>22.2</td>
<td>208</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>41.9</td>
<td>67.5</td>
<td>169.2</td>
<td>52.1</td>
<td>23.4</td>
<td>198.3</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>4.13</td>
<td>5.01</td>
<td>3.03</td>
<td>3.37</td>
<td>1.15</td>
<td>3.03</td>
</tr>
</tbody>
</table>
Overall indices of predictive performance are given in Table 2. With all four models, bias (MDPE) was \(\leq 15\%\) and inaccuracy (MDAPE) \(<25\%\). No significant difference was seen between the models with respect to MDAPE but both the White and Schuttler models showed significantly more underprediction than Diprifusor. No significant difference occurred between the models with respect to MDAPE but both the White and Schuttler models showed significantly more underprediction than Diprifusor. No significant difference occurred between the models in terms of wobble or divergence based on APE. However, when divergence was determined using signed PE, both the White and Schnider models differed significantly from Diprifusor. There was no significant correlation with any of the models between the duration of the maintenance infusion and the degree of divergence seen in each patient.

Table 3 gives MDPE values based on samples collected at different phases of the study. During the rapid infusion of propofol (21 mg kg h\(^{-1}\)) a positive bias was seen with Diprifusor but a negative bias occurred with the other three models. In the early phase from 1–21 min, similar differences between the models were observed but only with the White model was the difference significant. In the maintenance phase, all models showed some positive bias, with the White and Schuttler models demonstrating a significant improvement over Diprifusor. In the recovery phase, the negative bias seen with the Schuttler model was significantly greater than that with Diprifusor. In this phase the Schneider model showed a positive bias, which was significantly different from Diprifusor. There was no significant correlation with any of the models between the duration of the maintenance infusion and the overall or maintenance phase MDPE. Table 4 provides information on the number of patients at each phase in which MDPE was within the range \(\pm 20\%\) out of the total of nine studied. The differences seen were not statistically significant.

The mean measured propofol concentration (SD) peaked at 4.69 (0.51) mg ml\(^{-1}\) at 5 min and at 120 min was 3.25 (0.31) mg ml\(^{-1}\). To illustrate some of the differences in predictive performance shown in Tables 2 and 3 and to link these with differences between the models in input parameters and associated volumes and clearances, a further

![Fig 1 Box and whisker plot showing inter-patient variability in measured propofol concentrations in the nine patients studied. Maximum, minimum, median, and 25th and 75th percentiles are shown at selected time points.](https://academic.oup.com/bja/article-abstract/102/5/626/253153/628)

Table 2 Overall indices of predictive performance (medians and ranges). Significantly different from Diprifusor group. \(*P<0.05; **P<0.005\). PE, performance error; MDPE, median PE; MDAPE, median absolute PE.

<table>
<thead>
<tr>
<th>Model</th>
<th>MDPE (%)</th>
<th>MDAPE (%)</th>
<th>Divergence APE (% h(^{-1}))</th>
<th>Divergence PE (% h(^{-1}))</th>
<th>Wobble (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diprifusor</td>
<td>2.3 (−31.6 to 33)</td>
<td>24.6 (11.2–37.3)</td>
<td>−2 (−9.2 to 8.6)</td>
<td>−2.4 (−22.2 to 12.9)</td>
<td>18.8 (11.8–23.9)</td>
</tr>
<tr>
<td>White</td>
<td>−12.6 (−32.8 to 16.5)*</td>
<td>21.4 (13.2–37.2)</td>
<td>−1.4 (−11.2 to 11)</td>
<td>1.2 (−14.4 to 19.7)**</td>
<td>17 (10.8–25.8)</td>
</tr>
<tr>
<td>Schuttler</td>
<td>−6.2 (−32.4 to 17.9)*</td>
<td>20.6 (8.4–33.3)</td>
<td>−1.2 (−4.9 to 12)</td>
<td>−2.3 (−19.8 to 12.4)</td>
<td>13.5 (11.1–24.5)</td>
</tr>
<tr>
<td>Schnider</td>
<td>−0.1 (−21.5 to 33.5)</td>
<td>23.6 (13.1–42.8)</td>
<td>1.4 (−5 to 14.8)</td>
<td>13.2 (−6 to 33)**</td>
<td>18.8 (12.3–46.3)</td>
</tr>
</tbody>
</table>

Table 3 Median performance error % (median and range) based on samples collected at different phases of the study. Significantly different from Diprifusor group. \(*P<0.05; **P<0.005\).

<table>
<thead>
<tr>
<th>Model</th>
<th>Rapid infusion (1–5 min)</th>
<th>Early phase (1–21 min)</th>
<th>Maintenance (25 min – end-infusion)</th>
<th>Recovery (2–20 min post-infusion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diprifusor</td>
<td>17.7 (−41.6 to 71.1)</td>
<td>7.2 (−43.2 to 21.7)</td>
<td>12.7 (−28.7 to 53)</td>
<td>−10.5 (−46.1 to 64.6)</td>
</tr>
<tr>
<td>White</td>
<td>−14.9 (−53.1 to 27.7)**</td>
<td>−12.5 (−51.9 to 6.9)**</td>
<td>3 (−29.8 to 42.3)*</td>
<td>−8.7 (−46.7 to 64.7)</td>
</tr>
<tr>
<td>Schuttler</td>
<td>−21.4 (−53.1 to 21.9)**</td>
<td>−5.4 (−37.7 to 18.1)</td>
<td>5.3 (−28.1 to 40.9)*</td>
<td>−25 (−52.4 to 20.6)**</td>
</tr>
<tr>
<td>Schnider</td>
<td>−36.8 (−59.9 to 5.7)**</td>
<td>−20 (−35.9 to 15.7)</td>
<td>9.9 (−15.6 to 63)</td>
<td>15.5 (−25.8 to 96)**</td>
</tr>
</tbody>
</table>
simulation of a 2-h infusion and 20 min recovery period was done with parameters for a 70 kg, 170 cm, 50-yr-old male patient (Table 5). Results are shown together with mean measured propofol concentrations in Figure 2A–C.

During the rapid infusion (0–5 min; Fig. 2A), Diprifusor underpredicts the mean measured concentrations while these are overpredicted to a small extent by the White and Schuttler models but to a much larger extent by the Schnider model. As the infusion continues, the Diprifusor and Schuttler models show the smallest prediction error while the White and Schnider models continue to overpredict. During the ‘maintenance’ phase (Fig. 2B), Diprifusor continues to underpredict measured concentrations while the other three models follow the measured values more closely. In the recovery phase (Fig. 2c), both the Diprifusor and White models follow the measured profile closely while the Schnider model underpredicts the first few measured concentrations and the Schuttler model overpredicts concentrations in the later part of this phase.

**Discussion**

The simulated profile used in the present study differs from a TCI infusion where an initial loading dose is usually delivered at a rate of 1200 ml h$^{-1}$, equivalent to about 170 mg kg$^{-1}$ h$^{-1}$ for a 70 kg patient. However, previous results obtained in a simulation study with this infusion data and the Marsh, Dyck and Shafer, and Tackley models were in good agreement with results subsequently obtained with these same models in a clinical TCI study and the maintenance and recovery phases in the present study closely resemble the situation with TCI. Positive bias (MDPE) occurs when the measured concentration exceeds the predicted concentration, that is the model has underpredicted the measured concentration. Thus the terms ‘positive bias’ and ‘underprediction’ mean the same thing. The clinical consequence of using a model that underpredicts for TCI, is that the actual blood concentration achieved may be greater than the value indicated by the pump. Opposite effects occur with a model which overpredicts the measured concentration and shows negative bias.
Based on the overall indices of predictive performance (Table 2), with divergence assessed on the basis of APE, all four of the pharmacokinetic models provided performance values similar to those that have been deemed clinically acceptable in earlier studies. However, the overall values mask some differences between the models that become apparent when samples collected at different phases of the infusion are examined (Table 3). With the Schnider model, the overall figure of −0.1% for MDPE indicates negligible bias but is achieved as a consequence of overprediction (predicted values greater than measured) in the early phase being countered by underprediction in the recovery phase. Opposite effects are seen with Diprifusor with underprediction during rapid infusion and overprediction in the recovery phase. The White and Schuttler models were more likely to overlap at both early and late phases but both showed minimal bias during the maintenance phase.

When assessed on the basis of changes in signed PE over time, significant positive divergence was seen with the Schuttler model as a consequence of the changing bias seen with this model between early and late phases. This approach appears to be more informative than the conventional assessment of divergence based on APE, as the latter may be misleading if the direction of the error changes over time.

The difference between the Diprifusor and White models can be attributed to the smaller V1 in the White model as the other parameters are unchanged apart from a small reduction in metabolic clearance. It is likely that the overprediction seen in the first 5 min with the Schnider model is also related to the smaller V1 in this model but with a continuous infusion the height of the initial peak will also be influenced by k12 and k10. The overprediction noted in the recovery phase with the Schuttler model is consistent with lower clearance provided by this model.

With the small number of patients studied and the limited range of patient ages (24–56 yr) and BMI (19.5–29.4 kg m⁻²) there is no clear evidence of a marked improvement in predictive performance with the models incorporating patient covariates. The number of patients with MDPE values within the range of ±20% was similar for the overall assessment and in the early phase. In the maintenance phase, more patients met this criterion with the covariate models relative to Diprifusor, but with the Schuttler and Schnider models, this was accompanied by poorer predictive performance in the rapid infusion and recovery phases. Nevertheless, specifically in frail, elderly, and both patients, overprediction during the induction phase will not convey any risk of overdose, whereas underprediction may lead to excessive dosage. Changes in propofol pharmacokinetics described in elderly patients include a reduction in initial volume of distribution and clearance, a reduction in rapid peripheral distribution, and sex-related differences in volumes of distribution and clearance, all leading to increased concentrations if the same dose as in younger patients is administered. As a consequence, it is highly probable that a greater benefit of the covariate models would be seen in a population including elderly patients. As far as obese patients are concerned, propofol pharmacokinetic parameters appear to be scaled to total body weight and in this context, the Marsh model gives satisfactory bias and inaccuracy.

The underprediction of measured values seen at most phases in this simulation study with the Diprifusor model is consistent with the observation of a positive bias around or <20% in most of the studies that have investigated the predictive performance of Diprifusor TCI when used for anaesthesia although one study noted a positive bias of much greater magnitude. Other studies have reported a small negative bias overall (−1.4%, −5.3%, and −12.1%, respectively). Two of these studies describe an increase in the negative bias when the concentration is decreasing, and an increasing positive bias at higher target concentrations has also been noted. In the use of Diprifusor TCI systems for sedation, negative bias of −12% and −47% was described in two studies. A further study reported no overall bias but precision assessed by regression analysis was considered to be poor.

In intensive care unit (ICU) patients sedated with Diprifusor TCI, with target concentrations in the range of 0.2–2.0 μg ml⁻¹, the median MDPE was 17.9% in post-cardiac surgery patients and −6.6% in 10 general ICU patients. In 10 critically ill hypoalbuminaemic patients, median MDPE was −7% was not significantly different from the value of −2% in normoalbuminaemic patients.

Thus the majority of studies with Diprifusor TCI have demonstrated a level of predictive performance that has been considered clinically acceptable in a range of clinical situations and in association with different analgesic supplements. The overall trend is for measured values to exceed the target during anaesthesia with slight overprediction (negative bias) during sedation or recovery from anaesthesia. The overprediction noted with the White model in the early phase of this simulation study relative to Diprifusor would lead to reduced drug delivery with the new model if used for TCI and would be expected to reduce the positive bias usually seen with Diprifusor systems. The Schuttler model was derived from a population analysis of data provided by five research groups with different modes of administration, differing sampling sites, and a wide range of ages and weights and we are not aware of any prospective validation of this model.

The Schnider model was developed from a study in 24 volunteers given a constant infusion of propofol of 25, 50, 100, or 200 μg kg⁻¹ min⁻¹ over 60 min. In this model, V1 is constant at 4.27 litres for all patients, V2 and Cl2 increase if age >53 yr and decrease if >53 yr. Metabolic clearance, Cl1 is related to weight, height, and lean body mass (LBM) such that it increases if LBM <59 kg and decreases if LBM >59 kg. The only information on the prospective validation of this model when used for TCI appears to be in volunteer studies. Doufas and colleagues targeted effect site concentrations suitable for sedation and collected venous
samples 9 and 15 min after each increment in target-setting. MDPE was −13% when propofol was supplemented with N₂O and −18% when patients breathed air. This negative bias may be explained by venous sampling. Samples were obtained to demonstrate the achievement of pseudo steady-state and no early or recovery samples were collected. A second study by this group, in 18 volunteers, also targeted effect site concentrations, which were increased at different rates until various clinical endpoints, including recovery of responsiveness, were reached. Predictive performance assessed using arterial samples demonstrated minimal bias (1.75%) and good accuracy (MDAPE of 21%). Struys and colleagues showed that arterial propofol concentrations, measured over 5 min after a bolus dose of 2.5 mg kg⁻¹ given over 10 s, were poorly predicted by both the Marsh and Schnider models.

An ideal model should perform well at all stages of an infusion and this study has shown that differences exist between the four models studied, which would not be detected by a conventional assessment of predictive performance as advocated by Varvel and colleagues. On the basis of the results in Tables 3 and 4, the White model showed a marginal improvement over Diprifusor and would be expected to reduce the positive bias usually seen with any model is limited by marked inter-patient pharmacokinetic variability leading to a wide range of measured propofol concentrations despite the standardized infusion regimen used in this study. The method of calculating divergence is not always clearly stated in published studies. We propose that the calculation of divergence should involve linear regression analysis of absolute and signed PEs as both provide useful information. Further studies of this type in patients with a wider range of age, body weight, and body mass index would be desirable.

Acknowledgement

We are grateful to Mr Nick Syrett of AstraZeneca for his advice on statistical methodology.

References

3 Glen JB. The development of ‘Diprifusor’: a TCI system for propofol. Anaesthesia 1998; 53: 13–21
26 Pandin PC, Cantraine F, Ewalenko P, Deneu SC, Courxaert E, d’Hollander AA. Predictive accuracy of target-controlled propofol
and sufentanil coinfusion in long lasting surgery. *Anesthesiology* 2000; 93: 653–61


33 Doufas AG, Bakhshandeh M, Bjorksten AR, Shafer SL, Sessler DI. Induction speed is not a determinant of propofol pharmacodynamics. *Anesthesiology* 2004; 101: 1112–21