Comparing patient generated blood glucose diary records with meter memory in type 2 diabetes

J.E. Given a,*, M.J. O’Kane b, V.E. Coates a,c, A. Moore d, B.P. Bunting e

a Institute of Nursing and Health Research, University of Ulster, Coleraine Campus, Cromore Road, Coleraine, County Londonderry BT52 1SA, UK
b Department of Clinical Chemistry, Western Health and Social Care Trust, Altnagelvin Hospital, Londonderry BT47 6SB, UK
c School of Environmental Sciences, University of Ulster, Coleraine Campus, Cromore Road, Coleraine, County Londonderry BT52 1SA, UK
d Nursing Directorate, Western Health and Social Care Trust, Altnagelvin Hospital, Londonderry BT47 6SB, UK
e School of Psychology, University of Ulster, Magee Campus, Londonderry BT48 7JL, UK

A R T I C L E   I N F O
Article history:
Received 26 April 2013
Received in revised form
24 February 2014
Accepted 4 March 2014
Available online 11 March 2014

Keywords:
Blood glucose
Type 2 diabetes
Measurement
Validity
Medical records

A B S T R A C T
Aim: To assess agreement between meter and diary self monitoring of blood glucose (SMBG) records, over a year, in a sample of patients with type 2 diabetes.

Methods: Meter and diary records were available, for 95 individuals, who took part in the Efficacy of self monitoring of blood glucose in patients with newly diagnosed type 2 diabetes study.

Pearson’s correlation coefficient was used to explore the relationships between the types of error. Maximum likelihood estimation was used to explore changes over time through a structural equation modelling approach. Paired samples t-tests were used to determine if the presence of errors led to a significant difference between the mean diary and meter SMBG concentrations or coefficients of variation. Multiple regression was used to explore possible predictors of the error indices.

Results: Mean over-reporting, under-reporting, concordance and overall reliability were 8.4%, 10.0%, 83.5% and 71.3%, respectively. The first week of monitoring had significantly more under-reporting, over-reporting and less concordance and overall reliability than subsequent weeks. The majority of concordance errors were not clinically significant. Those that were, tended to occur during the first three months of monitoring. Participants’ at one trial site were significantly more likely to have recording errors than those at the largest site. Conclusions: Error levels were similar to those described previously in type 1 diabetes and there was a suggestion of an initial learning curve for record keeping. For some individuals diary records would not be considered acceptable if held to the same standards as blood glucose meters.

© 2014 Elsevier Ireland Ltd. All rights reserved.

* Corresponding author. Tel.: +44 2870124094; fax: +44 2870124951.
E-mail address: Given-J@email.ulster.ac.uk (J.E. Given).

Abbreviations: SMBG, self monitoring blood glucose; ESMON, efficacy of self monitoring of blood glucose in patients with newly diagnosed type 2 diabetes study; ANOVA, analysis of variance; NIMDM, Northern Ireland Multiple Deprivation Measure.
http://dx.doi.org/10.1016/j.diabres.2014.03.003
0168-8227/© 2014 Elsevier Ireland Ltd. All rights reserved.
1. **Introduction**

Most modern blood glucose meters can record the value, date and time of each blood glucose measurement [1]. Despite this clinicians still rely heavily on patient diaries when it comes to making therapeutic decisions based on self monitoring of blood glucose (SMBG) data [2]. This may be because meter memory reviews take more time, are more costly and can be complicated and inconvenient for health care providers to perform [1,3]. However, diaries consist of self-reported glucose values and may contain intentional or unintentional recording ‘errors’.

A recent systematic review, exploring diary and meter SMBG record agreement, identified three types of recording error in patient diaries. These were incorrectly recording a value that has been measured (lack of concordance), failing to record a value that has been measured (under-reporting) and adding a value to the diary that has not been measured (over-reporting) [4]. Allowing for a minimal amount of disagreement between the meter and diary SMBG records just over 50% of adult diaries could be considered as ‘accurate/reliable’ [4,5].

However, little information was available specifically relating to those with type 2 diabetes, perhaps because monitoring was not common practice in this population when the majority of articles were published [4,5]. This is an important gap in the literature considering that, in developed countries, the majority (85–95%) of those with diabetes have type 2 diabetes [6,7]. Meter and diary agreement were also assessed over limited time frames [3,8,9] ignoring the potential for changes in recording behaviour with time.

The aim of this study was to explore the level of agreement between diary and meter SMBG records, over a year, in a sample with type 2 diabetes.

2. **Subjects**

Blood glucose meter and diary SMBG records were available for 95 individuals, 55 men and 40 women ranging from 35 to 80 years of age (mean 57.6), who had taken part in the Efficacy of self monitoring of blood glucose in patients with newly diagnosed type 2 diabetes (ESMON) study [10]. This was a multicentre, year long, randomised controlled trial which aimed to assess the effect of SMBG on glycaemic control and psychological indices in patients with newly diagnosed type 2 diabetes. All patients were treated with lifestyle intervention and oral hypoglycaemic agents as necessary, according to a defined treatment algorithm. Ethical approval for the ESMON study was obtained from the University of Ulster ethics committee. Those in the monitoring group were asked to monitor their blood glucose eight times a week. At each three month review appointment, as part of the trial protocol, their meter SMBG record was printed and their diaries were collected.

3. **Methods**

It was suspected that monitoring behaviour and the amount of agreement may vary in relation to clinic appointments. As a result, for each three month review period the first week after clinic, the middle week between clinics and the week directly before clinic were explored. Over these 12 weeks percentage under/over-reporting, concordance and overall reliability, which has been proposed as a combined measure of agreement [11], were determined. These were calculated as follows:

\[
\% \text{ Over-reporting} = \left( \frac{\text{number of values added to diary/number of diary values}}{100} \right)
\]

\[
\% \text{ Under-reporting} = \left( \frac{\text{number of values not recorded in diary/number of meter values}}{100} \right)
\]

\[
\% \text{ Concordance} = \left( \frac{\text{number of meter values recorded accurately/number of meter values recorded at all}}{100} \right)
\]

\[
\% \text{ Overall Reliability} = \left( \frac{\text{number of meter values recorded accurately/number of meter values + number of values added to diary}}{100} \right)
\]

Other words this is the number of times the meter was used and readings correctly entered into the diary, as a percentage of the total number of times over-reporting, under-reporting, concordant or non-concordant reporting occurred.

The clinical significance of concordance errors was determined using the Clarke error-grid analysis. This is an established tool for assessing blood glucose meter accuracy when compared with laboratory measurement [12]. It determines the clinical significance of the difference between two blood glucose estimations by taking into account the size of the difference and the absolute blood glucose concentration [13]. This is important as two measurements which differ by the same amount from their true values (e.g. 50%) can have radically different clinical implications depending on the patient’s true blood glucose level [14].

The mean SMBG value and the coefficient of variation were also determined for the diary and meter at each time point.

The exploration of any change over time in the error indices was complicated by missing data, one of the most pervasive problems in data analysis [15]. In order to have complete data patients must remember to bring both meter and diary to every clinic appointment and measure and record blood glucose values at least once each week. The most popular method of dealing with missing data is listwise deletion. This means that all cases which have a missing value for any of the variables in the data are excluded from all computations. In comparison in pairwise deletion only cases with missing values on variables tagged for a particular computation are excluded from the analysis [16]. Both these approaches result in a decrease in sample size with a resulting decrease in statistical power. For example in this sample listwise deletion would have significantly reduced the sample size for all the error indices, from \( n = 24 \) for concordance to \( n = 38 \) for overall reliability.

An alternative theory based approach to missing data is full information Maximum Likelihood estimation [16]. Full information Maximum Likelihood estimation uses all the available data while reducing the bias and decrease in statistical power that are associated with listwise or pairwise deletion [16,17]. It was used in a structural equation modelling framework, a statistical methodology that takes a hypothesis-testing approach to the analysis of relationships between variables [18]. Structural equation modelling allows for greater flexibility than traditional multivariate methods when testing statistical
restrictions on the analysis of means, variances and covariances. With traditional multivariate methods there is the assumption of sphericity, that is variances and covariances are essentially equal across time [19]. Structural equation modelling allows relaxation of this assumption, an important consideration given that it was evident from the data that the variances (square of the standard deviations) were not equal across time, see Table 1. IBM SPSS Statistics AMOS (Meadville, USA) version 20 was used to model the percentage of each type of error over time. A one-way repeated measures analysis of variance (ANOVA), and Friedman’s test where appropriate, were also performed.

Pearson’s correlation coefficient was used to explore the relationship between the average error indices as well as their relationship with the total number of monitoring episodes. Paired samples t-tests were used to determine if recording errors led to a significant difference between the mean diary and meter SMBG concentration or coefficients of variation. Multiple regression was performed to determine if any of the error indices were predicted by age, gender, hospital site or Northern Ireland Multiple Deprivation Measure 2005 (NIMDM) score. The NIMDM is the official measure of area based multiple deprivation recommended for use by the Northern Ireland Government. It includes measures of income deprivation, employment deprivation, health deprivation and disability, education, skills and training deprivation, proximity to services, living environment and crime and disorder [20]. The NIMDM is calculated at Census Output Area level which is the lowest geographical scale for which census data is provided in the United Kingdom (with an average population size of about 300 in Northern Ireland).

4. Results

For the population as a whole over the twelve weeks, 6596 blood glucose readings were measured and 6829 values recorded in a diary. On average subjects performed 6.9 (range 1.2–11.4, SD 2.0) blood glucose tests each week and recorded 7.4 (range 0.4–26.0, SD 3.4) values in their diary. Not all individuals had both a meter and a diary available for verification at each time point but for those who did concordance was assessed, along with the clinical significance of any errors, for 5708 values.

In this sample of those with type 2 diabetes, over the year, the average under-reporting was 10.0%, over-reporting was 8.4%, concordance was 83.5% and overall reliability was 71.3%. In terms of individual diaries, the ranges for under/over-reporting and concordance show that the ESMON population has individuals at both extremes of record keeping. While some individuals had diaries that were almost a complete match to their meter record, others had diaries which were almost completely erroneous, see Table 1.

Out of a total 868 concordance errors only 16 were clinically significant and 76 potentially clinically significant as determined using the Clarke Error-Grid Analysis. Seven individuals were responsible for all the clinically significant errors, more than half of which occurred in the first three-month period. Clinically significant errors were not always an attempt to improve the diary impression of glycaemia as about one third increased the corresponding meter value.

When attempting to determine if there was any change in the error indices over time, the data did not support models where the amount of error was consistent over all 12 weeks. Instead, the first week of over-reporting was much greater than subsequent weeks and relaxing the restriction, for this first week to be equal to the others, produced a model with good fit, $\chi^2 (10) = 11.77$, $p = 0.30$, TLI 0.94, CFI 0.99. The same model also held for concordance, $\chi^2 (2) = 4.96$, $p = 0.08$, TLI 1.13, CFI 1.0, and overall reliability, $\chi^2 (10) = 9.73$, $p = 0.46$, TLI 1.01, CFI 1.0. This indicates that over-reporting, concordance and overall reliability were significantly worse during the first week of the study. While Friedman’s test and ANOVA did not find the first week of monitoring to be significantly worse than subsequent weeks this is unsurprising given the reduced sample size and associated loss in statistical power associated with these tests.

For under-reporting a statistically significant difference was found between weeks 1 and 7 and 27 and 33 when compared with the remaining weeks, which were restricted to have equal means (within sampling error), $\chi^2 (7) = 7.44$, $p = 0.39$, TLI 0.99, CFI 1.0. Friedman’s test found the same results for under-reporting with an additional difference in weeks 33 and 39 suggested by ANOVA ($p < 0.05$).
Over the year, there was a moderate positive correlation between over-reporting and under-reporting, \( r = 0.56, n = 89, p < 0.001 \), indicating that as over-reporting increased so did under-reporting. This has been described previously [11,21,22]. Over-reporting was moderately negatively correlated with concordance, \( r = -0.32, n = 88, p < 0.01 \) indicating that as over-reporting increased concordance decreased. Under-reporting and concordance were also significantly negatively correlated, \( r = -0.33, n = 88, p < 0.01 \) indicating that as under-reporting increased concordance decreased. Overall reliability was strongly correlated with the other error indices, \( r = -0.715 \) to \(-0.775, p < 0.001\), supporting its use as summary index of meter and diary agreement.

There was no significant correlation between the total number of blood glucose values measured (using the meter) and average over-reporting or concordance. There was a small negative correlation between the total number of SMBG values measured and under-reporting, \( r = -0.275, n = 89, p < 0.01 \), indicating that as the total number of SMBG values measured increased the amount of under-reporting decreased. There was also a small positive correlation between the total number of SMBG values measured and overall reliability, \( r = 0.214, n = 89, p < 0.05 \), indicating that as patients measured more blood glucose values they had greater overall reliability.

There was no statistically significant difference between the mean meter and diary value or coefficient of variation over the year.

Average over-reporting was not predicted by age, gender, hospital attended or the NIMDM 2005 score of the area within which the individual lived. Age, gender and the NIMDM 2005 score of the census output area within which the individual lived did not predict average under-reporting, concordance or overall reliability. At one ESMON study site patients were statistically significantly more likely to score higher on under-reporting (B = 12.31, S.E. = 4.47), and lower in concordance (B = -9.18, S.E. = 3.87) and overall reliability (B = -18.04, S.E. = 5.47) than those at the largest site. Whether this is due to a difference in clinical care between hospitals or another reason is unclear.

5. Discussion

It is possible that the participants in the ESMON trial were more motivated and therefore not fully representative of the general population [23]. If so, then the error described here could be taken as the minimal amount that would be expected in a population with type 2 diabetes. Over-reporting in this cohort was less than half that reported previously in patients with type 2 diabetes [24]. Indeed it was much more in line with that described in patients with type 1 diabetes [9,25,26]. Under-reporting, concordance and overall reliability were also very similar to that described in type 1 diabetes [4].

Although infrequent, clinically significant errors may potentially lead to either inappropriate lowering of blood glucose or failure to treat high blood glucose if the diary value was taken as correct. Using the Clarke error-grid analysis to determine the clinical significance of concordance errors was a new approach but it was felt to be important as the clinical significance of an ‘error’ is not solely dependent on its size [14]. If patient diaries were to be held to the same standards of accuracy as blood glucose meters, then at least 95% of tests should be correct or not clinically significantly different and none should be clinically significantly different [27]. In ESMON 98.4% of meter values met these criteria. For seven subjects the presence of clinically significant errors meant that diary records were not considered acceptable under these standards.

There was a suggestion of an early learning curve with significantly more under/over-reporting and less concordance and overall reliability during the first week of record keeping compared with subsequent weeks. This is supported by the fact that the majority of clinically significant concordance errors occurred in the first three month period and would suggest that initial early SMBG records should be used with caution when making clinical decisions. The additional differences for under-reporting may suggest that there is more under-reporting in the week immediately following clinic and that it improves by the week midway between appointments but more work would be needed to clarify this.

Overall reliability was strongly correlated with the other error indices supporting its use as a combined measure of meter and diary agreement. However, it would be difficult to calculate overall reliability in the limited time allowed in the clinical setting. Instead over-reporting, which is moderately correlated to both under-reporting and concordance, could be used as a simple measure of meter and diary agreement in the clinical setting. It would be much more feasible to quickly assess the number of values added to the diary than to compare all the diary and meter values in the level of detail required to spot concordance errors etc.

6. Conclusion

Through this study it has been possible to describe the amount of under-reporting, concordance and overall reliability in the blood glucose diaries of a sample of patients with type 2 diabetes. This information was not previously available and is important as there has been a general increase in monitoring among those with type 2 diabetes [28,29]. In terms of recording errors there was a suggestion of an initial learning curve followed by error levels similar to those described in patients with type 1 diabetes.

Conflict of interest

MJO’K has received research funding from Nova Biomedical, a manufacturer of blood glucose meters.

Acknowledgements

The efficacy of self monitoring of blood glucose in patients with newly diagnosed type 2 diabetes (ESMON study) was funded by the Northern Ireland Research and Development Office, Belfast.
This work will constitute part of a PhD for JG which has been funded by the Department for Employment and Learning for Northern Ireland.

There was no involvement of either funder in study design, data collection, data analysis and manuscript preparation or publication decisions.

REFERENCES