Original Article

Journal of Diabetes Science and Technology
1–8

DOI: 10.1177/19322968241264744
journals.sagepub.com/home/dst
Sage

Carbohydrate Estimation Accuracy of Two Commercially Available Smartphone Applications vs Estimation by Individuals With Type 1 Diabetes: A Comparative Study

Michelle Baumgartner, MSc[1,2], Christian Kuhn, MSc[1], Christos T. Nakas, PhD[3,4], David Herzig, PhD[1]*, and Lia Bally, MD, PhD[1]*

Abstract

Background: Despite remarkable progress in diabetes technology, most systems still require estimating meal carbohydrate (CHO) content for meal-time insulin delivery. Emerging smartphone applications may obviate this need, but performance data in relation to patient estimates remain scarce.

Objective: The objective is to assess the accuracy of two commercial CHO estimation applications, SNAQ and Calorie Mama, and compare their performance with the estimation accuracy of people with type 1 diabetes (T1D).

Methods: Carbohydrate estimates of 53 individuals with T1D (aged $\geq$16 years) were compared with those of SNAQ (food recognition + quantification) and Calorie Mama (food recognition + adjustable standard portion size). Twenty-six cooked meals were prepared at the hospital kitchen. Each participant estimated the CHO content of two meals in three different sizes without assistance. Participants then used SNAQ for CHO quantification in one meal and Calorie Mama for the other (all three sizes). Accuracy was the estimate's deviation from ground-truth CHO content (weight multiplied by nutritional facts from recipe database). Furthermore, the applications were rated using the Mars-G questionnaire.

Results: Participants' mean $\pm$ standard deviation (SD) absolute error was 21 $\pm$ 21.5 g (71 $\pm$ 72.7%). Calorie Mama had a mean absolute error of 24 $\pm$ 36.5 g (81.2 $\pm$ 123.4%). With a mean absolute error of 13.1 $\pm$ 11.3 g (44.3 $\pm$ 38.2%), SNAQ outperformed the estimation accuracy of patients and Calorie Mama (both $P > .05$). Error consistency (quantified by the within-participant SD) did not significantly differ between the methods.

Conclusions: SNAQ may provide effective CHO estimation support for people with T1D, particularly those with large or inconsistent CHO estimation errors. Its impact on glucose control remains to be evaluated.

Keywords
carbohydrate counting, decision support, diabetes, mHealth, smartphone application

[1]Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
[2]Department of Health Sciences and Technology, Eidgenössische Technische Hochschule Zurich, Zurich, Switzerland
[3]School of Agricultural Sciences, Laboratory of Biometry, University of Thessaly, Volos, Greece
[4]University Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

*Equally contributing last authorship.

Corresponding Author:
Lia Bally, MD, PhD, Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital, Bern University Hospital, University of Bern, Freiburgstrasse 15, Bern 3010, Switzerland.
Email: lia.bally@insel.ch

Introduction

Despite remarkable progress in diabetes technology, most treatment modalities still require estimating meal carbohydrate (CHO) content for meal-time insulin dosing.[1-3]

Accurate CHO quantification is not only burdensome[4] but also challenging and therefore prone to errors.[5,6] Innovative technologies enabling food analysis may support more accurate CHO quantification and reduce the burden of diabetes management.[4]

Image-based food recognition and quantification systems combining depth-sensing smartphone cameras with computer vision is an emerging field.[7,8] Although several commercial solutions, such as SNAQ or Calorie Mama, allow for automated food recognition,[9,10] the feature of volume estimation allowing for quantification of meal CHO content is currently restricted to very few apps.[9] Given the relevance of CHO determination for insulin dosing in people with type 1 diabetes (T1D),[11] insights into the accuracy of commercial food quantification systems and cross-comparison with estimation skills of patients are of interest to the patient community and caring health care professionals. Although several studies assessed CHO estimation skills of people with T1D[5,12] and accuracy of automated food recognition systems,[10,13] head-to-head (human vs device) comparisons allowing to rate technology performance relative to patient skills remain scarce.

In this study, we aimed to evaluate the accuracy of SNAQ and cross-compare it to the estimation performance of another app adopting a different technology (Calorie Mama), as well as to estimations by people with T1D.

Materials and Methods

Study Design and Procedures

The study was performed at the outpatient diabetes clinic of the University Hospital Bern, using meals served at the hospital canteen. Twenty-six meals (see Supplementary Appendices A and B) were chosen to create a representative sample of everyday cooked dishes. Each meal was prepared in three different sizes (1/4 portion, 1/2 portion, full portion), resulting in 78 different plates for evaluation. Ground-truth macronutrient content was determined by the research staff weighing each food item with a kitchen scale to the nearest 1 g. Nutrient information of the items was retrieved from the nutrition service software LogiMen (V5.4, LOGIMATIKA GmbH, Leonberg, Germany). LogiMen contains nutritional information from the German BLS and Swiss Food Composition nutrient databases. The Ethics Committee Bern was informed about the project and declared non-responsibility in the absence of health-related data collection (Req-2023-00230).

In a walk-in area, German-speaking individuals with T1D aged ≥16 years were offered to participate in the study after their routine medical consultation. Participants first estimated the CHO content of two meals (each served in three sizes) without any assistance, resulting in six estimates per participant. Thereafter, participants evaluated the CHO content of every plate with one of the two apps (Calorie Mama or SNAQ) in a randomized order. The first meal was evaluated by one of the two apps (in all three sizes), whereas the second was evaluated with the other app. Every participant was given a short instruction on the usage of the apps and had access to a user manual in German for both apps (Supplementary Appendix C). The automated food item recognition of the app was provided in German for SNAQ and English for Calorie Mama, as Calorie Mama is currently only available in English. If needed, a list containing the menu items in both languages (Supplementary Appendix A) was provided, to avoid any confounding by insufficient language skills.

A windowless environment with artificial light was chosen to minimize variability in lighting conditions. Afterwards, participants were asked to fill the German version of the Mobile Application Rating Scale (MARS-G, Supplementary Appendix D), evaluating the functionality and subjective quality of the apps.[14] The study procedure is visualized in Supplementary Appendix E.

Smartphone Apps

The apps SNAQ (Version 8.7.4, SNAQ GmbH, Zurich, Switzerland) and Calorie Mama (Version 5.330.41301, Azumio Inc, Palo Alto, California) were installed on an iPhone 12 Pro (Apple Inc, Cupertino, California) and were used according to their intended purpose. The rationale for the choice of test smartphone was the embedded depth sensor technology (currently only available on the newer iPhone models X, 11, and 12), allowing for volume estimation of recognized food items in the SNAQ app.

Both apps require the user to take a photo of the plate using the back camera (on iPhone models X and 11, SNAQ requires the use of the front camera to include depth-sensing technology). SNAQ proposes different food items on captured meal and requires to confirm the suggestions or select appropriate alternatives in a drop-down list. Once confirmed, depth-sensing technology allows for calculation of the food items' volume, which is subsequently converted into a weight using a food density database. Linkage with information of a nutrient database finally allows to assess the total macronutrient and energy content of the meal.[15] Conversely, Calorie Mama let the user confirm the recognized food items and calculates macronutrient quantities and energy content based on adjustable standard portion sizes. The sources of the incorporated bulk density values (SNAQ only) and nutrient databases of both applications are not disclosed by the manufacturers.

Target Metrics and Outcomes

The primary outcome was the absolute error of meal CHO content. The following target metrics were calculated to evaluate the accuracy of the CHO content estimates based on comparison with ground-truth (nutrient content as assessed using the measured weight and the nutrient information from the database) denoted as *reference*:

- Absolute error (units: [g] and [%])

The absolute estimation error is defined as: *Absolute error (%) = |estimation-reference|/reference and Absolute error (g) = |estimation – reference|*

- Bias (units: [g] and [%])

The bias of the estimation is defined as the difference between estimate and ground truth/reference (estimate – reference)

- Limits of agreement (units: [g] and [%])

The 95% limits of agreement were calculated as Bias $\pm$ 2 * SD (details regarding the calculation of the SD are reported in section "Statistical Analysis").

- Percentage of clinically relevant estimation errors (unit: [%])

Clinically relevant estimation errors were defined as absolute errors of >10 g of CHO in accordance with evidence on the impact of CHO counting errors on the quality of glycemic control.[16]

- CHO estimation consistency

The consistency of CHO estimation was calculated by the within-participant SD of the signed relative errors.

In addition to the accuracy in terms of meal CHO content estimation, the performance of the two apps regarding the estimation of energy, fat, and protein content was evaluated. Furthermore, a sensitivity analysis was conducted for the app SNAQ to identify the impact of different nutritional databases used between the app and the ground-truth calculation on the calculated errors (see section "Statistical Analysis"). The performance evaluation of the food recognition of the two apps was outside the scope of the study and is addressed elsewhere.[10] Usability was evaluated using the Mobile Application Rating Scale in German language (MARS-G).[14]

Statistical Analysis

All errors were determined on plate level (total CHO content of the plate). The Bland-Altman-type plots[17] were generated to visualize the level of agreement between the estimates and ground-truth values. Mixed-effect modeling[18,19] was used to estimate the accuracy metrics of interest, including standard deviations (SDs) (from which 95% Limits of Agreement [LoA] were calculated) as well as absolute and signed errors, according to the experimental procedure.[20] The ground-truth CHO content of the plate was treated as a fixed effect, whereas participants and meals were considered as random effects. In case of a significant difference between the methods, assessed using a Type III ANOVA with Satterthwaite's method, pairwise *P*-values were obtained by means of a Tukey's test using the emmeans package.[19] Data are described using mean $\pm$ SD, unless otherwise specified. All absolute errors and biases are presented as absolute values (g) followed by the relative values in parentheses. For the sensitivity analysis, performed to evaluate the influence of the different nutritional databases used by the apps and the ground-truth method, macronutrient content was calculated on the basis of the app's weight estimations (ignoring the embedded background nutrient database) and the corresponding nutritional information of the LogiMen software. Consequently, all the estimates were generated based on an identical nutritional database and were therefore only influenced by the accuracy of the app's volume quantification performance and food bulk density information. The sensitivity analysis was only applicable to SNAQ, as Calorie Mama does not allow exporting weight information of food items.

The consistency of the estimation method was evaluated by the respective within-participant SD of the apps and the participants. To account for possible influence by the study design (ie, the same participant rating the same meal in three different portion sizes could influence the consistency estimate), we also calculated the within-participant SD using only one (randomly selected) of the three portion sizes. The SD was compared between the methods using a linear model. The outcome variable (within participant-SD) was log-transformed to account for the non-normal distribution.

For the MARS-G questionnaire, mean and SD for each app and survey section were calculated. The internal consistency of the survey sections was evaluated using Cronbach's alpha. Survey scores were compared with a two-sided paired *t*-test. $P < .05$ were considered statistically significant. R version 4.1.2[21] was used for statistical analysis.

Results

In total, 26 meals in three different sizes were measured, resulting in 78 distinct plates. Fifty-three adults with T1D participated in the study, resulting in 318 patient estimations (six per participant) and 159 estimations using each app. The average meal energy content was 286 kcal (range = 73-827 kcal) and consisted of 29.0 g CHO (range = 6.9-100.1 g), 11.7 g fat (range = 1.7-49.2 g), and 14.4 g protein (range = 1.8-37.2 g). A photo of all study meals is included in Supplementary Appendix B.

Table 1. Carbohydrate Estimation Errors.

	Absolute error		Bias		LoA	Estimation error >10 g CHO
Method	Absolute value, g, mean (SD)	Relative value, %, mean (SD)	Absolute value, g, mean (SD)	Relative value, %, mean (SD)	Absolute value, g	Relative value, %
Participants	21.0 (21.5)	71 (72.7)	15.7 (16.2)	53.1 (54.8)	(-72.8, 104.2)	45.0
Calorie Mama	24.0 (36.5)	81.2 (123.4)	13.6 (16.6)	46 (56.1)	(-91.9, 119.1)	60.4
SNAQ	13.1 (11.3)	44.3 (38.2)	-1.8 (16.5)	-6.1 (55.8)	(-52.1, 48.5)	37.7

LoA: 95% limits of agreement; SD: standard deviation.

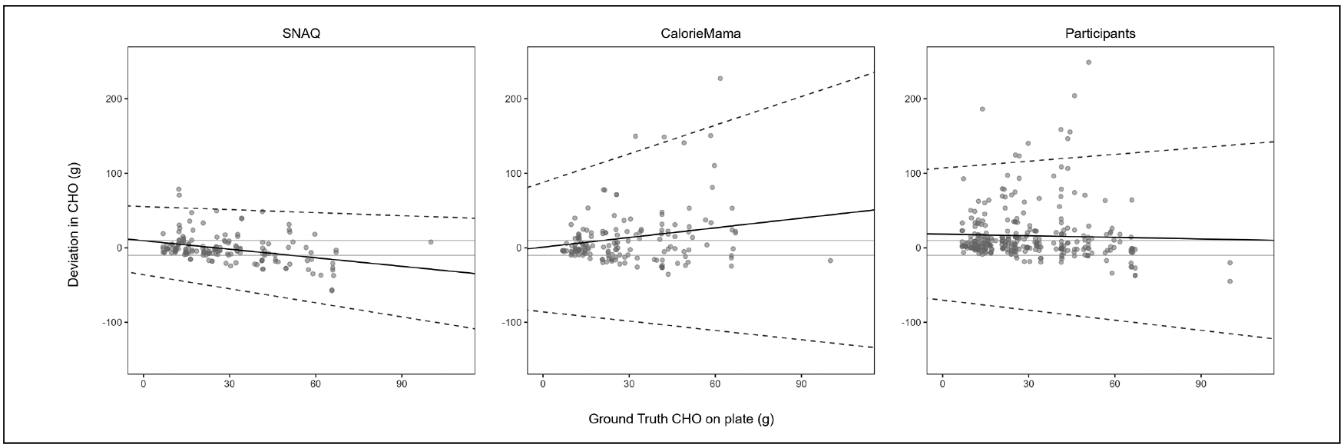

Figure 1. Bland-Altman plot illustrating the level of agreement between the estimated and reference carbohydrate content of the meals for each method. Bias (solid black line) and 95% limits of agreement (dashed lines) were determined using a linear mixed-effects model to account for the nested data structure. The gray solid lines represent the range for clinically relevant estimation errors (±10 g CHO).

Carbohydrate Content Estimation Accuracy

Results of all accuracy metrics are reported in Table 1. Bland-Altman-type plots for each method are shown in Figure 1. The absolute error of the patient estimation was 21.0 ± 21.5 g (95% confidence interval [CI] = 15.7 to 26.3) (71.0 ± 72.1%), whereas the absolute error of the estimation was 24.0 ± 36.5 g (95% CI = 17.8 to 30.2) (81.2 ± 123.4%) for Calorie Mama and 13.1 ± 11.3 g (95% CI = 6.9 to 19.3) (44.3 ± 38.2%) for SNAQ. The absolute estimation error by SNAQ was significantly lower compared with the participants' estimation error (P = .017) and the Calorie Mama estimation error (P = .003). The CHO estimation error of the patients and Calorie Mama did not differ significantly (P = .562). Although the errors of SNAQ were distributed around zero (Bias = −1.8 ± 16.5 g, [95% CI = −7.1 to 3.6]), patients and Calorie Mama tended to overestimate the CHO content (Bias = 15.7 ± 16.2 g [95% CI = 9.6 to 21.8] and 13.6 ± 16.6 g [95% CI = 5.2 to 22.0], respectively). Calorie Mama displayed the highest proportion of clinically relevant estimation errors (>10 g CHO) with 60.4%, followed by the patients with 45.0% and SNAQ with 37.7%.

Figure 2 illustrates the participant-specific SD in comparison to the mean SD of the two apps. Seventy-seven percent of the patient estimated more consistently than SNAQ and 83% estimated more consistently than Calorie Mama. There was no significant difference in the consistency of the estimation error between the patient and the two apps (P = .99). When randomly selecting one of the three meal sizes for the calculation of the within-participant SD (to account for possible influence of the study design), similar results were observed.

Estimation Accuracy for Energy, Protein, and Fat Content for Both Apps

A table summarizing the results of the estimation accuracy of energy, protein, and fat content for Calorie Mama and SNAQ can be found in Supplementary Appendix F. Calorie Mama yielded an absolute energy content error of 154 ± 157 kcal (53.9 ± 54.8%) and SNAQ of 133 ± 109 kcal (46.5 ± 38.2%). This difference in absolute error was not found to be statistically significant (P = .194).

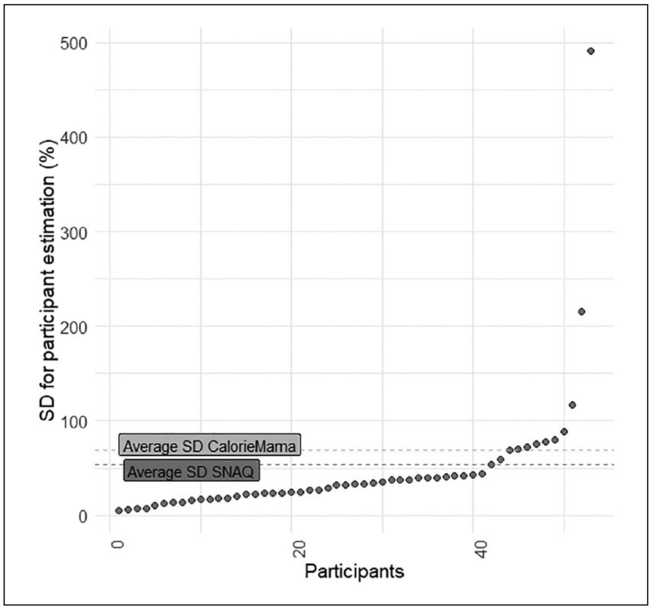

Figure 2. Within-participant SD of relative signed errors against the average SD of the two applications. Statistical analysis revealed that the difference in the mean SDs between the participants and the applications was not statistically significant.

Nutritional Database Sensitivity Analysis for SNAQ

As expected, the absolute error was reduced with the alignment of the nutritional databases (11.1 ± 9.1 g [38.8 ± 31.8%] vs 13.1 ± 11.3 g [44.3 ± 38.2%], P = .033). The detailed results of the sensitivity analysis for the accuracy of nutrient content estimation of SNAQ when using the identical nutritional database than the reference method (LogiMen) are illustrated in Supplementary Appendix G.

Usability Results

The MARS-G questionnaire was completed by 49 participants. Four participants were not able to independently operate the apps, either due to impairment of sight and/or dexterity, and were therefore not deemed eligible for usability rating. The questionnaire's section on subjective quality consists of four items with a Cronbach's α of 0.8, the functionality section likewise contained four items (α = 0.76). Both sections had a maximum possible score of five.

The mean score for the subjective quality was higher for SNAQ compared with Calorie Mama (3.6 ± 1.7 vs 2.7 ± 1.2, $P < .05$). Regarding the functionality of the apps, SNAQ received likewise higher scores compared with Calorie Mama (4.3 ± 0.7 vs 3.7 ± 0.4, $P < .05$). Results are visualized in Figure 3.

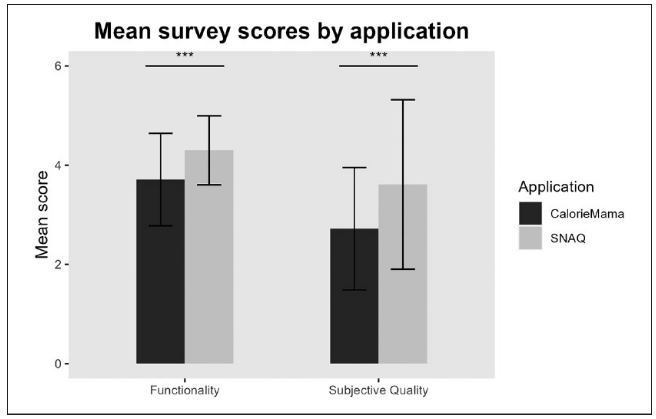

Figure 3. Mean functionality and subjective quality scores of the two applications assessed using the Mars-G questionnaire. Error bars represent the standard deviations.
*** indicates a P-value of less than 0.001.

Discussion

This study compared the CHO estimation skills of people with T1D against the CHO estimation accuracy of two commercial food analysis apps—Calorie Mama and SNAQ. SNAQ resulted in more accurate meal CHO content estimations than estimations by individuals with T1D, whereas Calorie Mama, requiring users to modulate a proposed standard portion size for quantification, did not outperform the estimation skills of people with T1D. The superiority of the estimation accuracy of SNAQ underscores that the magnitude of CHO error is critically influenced by the performance of food quantity estimation. However, although accuracy was significantly better, error consistency was not improved using SNAQ. This may be an essential point, as postprandial glucose control may largely depend on the consistency of the estimation errors rather than the overall accuracy of the estimates (ie, a constant relative deviation from the actual content could be compensated by adjusting carbohydrate-to-insulin factors).

In a previous study, where we evaluated the accuracy of SNAQ in terms of food quantity estimation,[9] we observed absolute errors in CHO of 14.8 ± 10.9%. This is considerably smaller than the absolute errors of 44.3 ± 38.2% observed in this study (or 38.8% when using the same nutritional database). However, it is important to note that in this earlier study, we tested SNAQ on test meals that were less complex than the meals used in this study. Another relevant design difference consists of the focus on food quantity estimation only (items that were wrongly detected were corrected by the study team) in the previous study, whereas in this study, the app was used according to real-life conditions. Although the relative absolute errors reported in the previous study, where slightly higher for cooked meals (18.2 ± 14.7%), values were still substantially smaller than the estimation error in the present evaluation. These findings reveal that the performance of SNAQ is critically dependent on the complexity of the meal (eg, multiple components, mounted above one another), the adequacy of the recognized items, the match of the nutritional database as well as the bulk density with actual food macronutrient composition and density. Consequently, the selection of test meals has a relevant impact on the performance assessment and should be critically appraised when drawing conclusions to real-life settings. In addition, food scenarios in real-life settings may be further complicated by variability in lighting conditions.

Evaluations of other image-based food analysis systems relying on artificial intelligence (AI)-empowered techniques have been described in the scientific literature. A study evaluating the accuracy of the commercial app Foodvisor, that estimates meal macronutrient content based on a food recognition algorithm and detection of food surface area derived from a single picture, reports a mean absolute estimation error of 13.9 ± 12.4 g (58 meals with an average CHO content of 94 g were tested).[22] In a study with a similar design as the present investigation, a research prototype allowing for both automated food recognition and quantification, was evaluated on the basis of 60 cooked meals obtained from a hospital restaurant and yielded a mean absolute error of 26.2 ± 18.7% (12.3 ± 9.6 g).[23] The same research prototype even generated better results in an earlier study with a reported CHO estimation error of 10% (6 ± 8 g).[24] However, unlike the SNAQ app evaluated in this study, the tested research prototype of these former studies requires the user to capture the meal from two different viewing angles, which may affect the usability of the system. Overall, significant differences in study design and experimental conditions (most notably the different complexity of the meals studied) make a fair comparison unrealistic.

Findings of the CHO estimation skills of people with T1D previously reported in the literature are variable, likely reflecting heterogeneity in study design and populations. Brazeau et al[5] assessed CHO estimation accuracy of adults with T1D on the basis of 72-hour food records and found a mean relative error of 20.9 ± 9.7%, corresponding to absolute values of 15.4±7.8 g. In our study, we found considerably higher relative errors (71 ± 72.7%). However, absolute errors were comparable (15.4 ± 7.8 g vs 15.7 ± 16.2 g), considering that the average meal CHO content was 29 g in this study compared with 72 g in the study by Brazeau et al. One may note here that an average meal CHO content of 29 g in this study is on the low side for a standard meal. In addition, the study participants in Brazeau et al estimated their everyday meals, which may be easier to estimate due to familiarity. The participants in our study were confronted with meals, which may not be typical of their usual diet. Another study evaluated the CHO estimation accuracy of 19 individuals with T1D by confronting them with a total of 60 different meals, resulting in 114 estimations.[23] The reported absolute error of 54.8 ± 72.3% (27.9±38.2.g) was comparable than the error observed in this study.

We acknowledge several limitations of this study. The average CHO content in the evaluated meals was low. Due to the inability to trace this data from both commercial apps, no information on the manual adjustments of the food recognition and quantity by the users and their relevance for the overall estimation error could be provided by this work. However, the main interest of this study was to report on the accuracy of commercial apps when used in line with real-life settings, which always reflects net results of a user-app interaction. All three meal sizes were presented simultaneously to the participants, potentially influencing their estimations. A sensitivity analysis was conducted to mitigate any impact on the consistency of the errors. However, we cannot completely rule out a possible effect on the magnitude of the errors. Nevertheless, we believe the impact on the overall results to be marginal. Finally, test meals may not be reflective of the participants' usual meal choices and portion sizes, thereby resulting in larger deviations of their CHO estimates.

Still, findings yielded by the present work provide important insights into the accuracy of two commercial image-based CHO estimation apps. Despite the relatively good accuracy of SNAQ compared with patients' estimations, the efficacy on glucose control, acceptability, and perceived benefits requires clinical evaluation. This is reinforced by the participants expressing hesitancy to use such an app on daily basis, mainly due to concerns of extra time investments. It should be mentioned that, the available evidence for the importance of accurate CHO quantification on glucose control is not overwhelmingly high (see Bell et al[25]), particularly in view of emerging automated insulin delivery systems which may provide forgiveness for imprecisions in CHO counting. Possible deployments outside everyday use may include its application for carbohydrate counting refresher courses and nutrition literacy training, optimization of carbohydrate-insulin and correction factors, management of unusual meals, and dietary intake assessments.

Conclusions

SNAQ may provide effective CHO estimation support for people with T1D, particularly those with large or inconsistent carbohydrate estimation errors. Its impact on glucose control remains to be evaluated.

Abbreviations

AID, automated insulin delivery; CHO, carbohydrate; LoA, limits of agreement; SD, standard deviation; T1D, type 1 diabetes.

Acknowledgments

The authors thank all the participants for their time. The authors specially thank the Insel Hospital Kitchen (Niklaus Fankhauser and

Beat Blum) for generously providing us with all the study meals and tableware. Furthermore, the authors show appreciation to the SNAQ team, specifically Aurelian Briner for the pleasant collaboration.

Author Contributions

MB, DH, CN, and LB contributed to conceptualization. MB and CK contributed to data curation. MB, CTN, DH, and CK contributed to formal analysis. MB contributed to investigation. MB, CTN, and DH contributed to methodology. MB contributed to project administration. LB contributed to resources. CTN, LB, and DH contributed to supervision. CTN and DH contributed to validation. MB contributed to visualization. MB contributed to writing—original draft. DH, CTN, and LB contributed to writing—review & editing. All authors approved the final draft.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: LB reports serving as a consultant for SNAQ.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was supported by an EFSD and EUDF Digitalisation in Diabetes Care Research Grant and by the Diabetes Center Berne. SNAQ offered two-month free SNAQ premium account for all the participants. SNAQ had no influence on the study including the design, interpretation of the results, and writing of the manuscript.

ORCID iD

Lia Bally https://orcid.org/0000-0003-1993-7672

Supplemental Material

Supplemental material for this article is available online.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Olczuk D, Priefer R. A history of continuous glucose monitors (CGMs) in self-monitoring of diabetes mellitus. *Diabetes Metab Syndr*. 2018;12(2):181-187. doi:10.1016/j.dsx.2017.09.005.
2. Beck RW, Bergenstal RM, Laffel LM, Pickup JC. Advances in technology for management of type 1 diabetes. *Lancet*. 2019;394(10205):1265-1273. doi:10.1016/S0140-6736(19)31142-0.
3. Hughes MS, Addala A, Buckingham B. Digital technology for diabetes. *N Engl J Med*. 2023;389(22):2076-2086. doi:10.1056/NEJMra2215899.
4. Fortin A, Rabasa-Lhoret R, Roy-Fleming A, et al. Practices, perceptions and expectations for carbohydrate counting in patients with type 1 diabetes—results from an online survey. *Diabetes Res Clin Pract*. 2017;126:214-221. doi:10.1016/j.diabres.2017.02.022.
5. Brazeau AS, Mircescu H, Desjardins K, et al. Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes. *Diabetes Res Clin Pract*. 2013;99(1):19-23. doi:10.1016/j.diabres.2012.10.024.
6. Reiterer F, Freckmann G. Advanced carbohydrate counting: an engineering perspective. *Annu Rev Control*. 2019;48:401-422. doi:10.1016/j.arcontrol.2019.06.003.
7. Lo FPW, Sun Y, Qiu J, Lo B. Image-based food classification and volume estimation for dietary assessment: a review. *IEEE J Biomed Health Inform*. 2020;24(7):1926-1939. doi:10.1109/JBHI.2020.2987943.
8. Dalakleidi KV, Papadelli M, Kapolos I, Papadimitriou K. Applying image-based food-recognition systems on dietary assessment: a systematic review. *Adv Nutr*. 2022;13(6):2590-2619. doi:10.1093/advances/nmac078.
9. Herzig D, Nakas CT, Stalder J, et al. Volumetric food quantification using computer vision on a depth-sensing smartphone: preclinical study. *JMIR Mhealth Uhealth*. 2020;8(3):e15294. doi:10.2196/15294.
10. Van Asbroeck S, Matthys C. Use of different food image recognition platforms in dietary assessment: comparison study. *JMIR Form Res*. 2020;4(12):e15602. doi:10.2196/15602.
11. Abreu C, Miranda F, Felgueiras P. Carbohydrate counting: how accurate should it be to achieve glycemic control in patients on intensive insulin regimens? *AIP Conf Proc*. 2019;2116(1):250009. doi:10.1063/1.5114249.
12. Gillingham MB, Li Z, Beck RW, et al. Assessing mealtime macronutrient content: patient perceptions versus expert analyses via a novel phone app. *Diabetes Technol Ther*. 2021;23(2):85-94. doi:10.1089/dia.2020.0357.
13. Sasaki Y, Sato K, Kobayashi S, Asakura K. Nutrient and food group prediction as orchestrated by an automated image recognition system in a smartphone app (CALO mama): validation study. *JMIR Form Res*. 2022;6(1):e31875. doi:10.2196/31875.
14. Messner E-M, Terhorst Y, Barke A, et al. The German Version of the Mobile App Rating Scale (MARS-G): development and validation study. *JMIR Mhealth Uhealth*. 2020;8(3):e14479. doi:10.2196/14479.
15. Swiss Food Composition Database. https://naehrwertdaten.ch/en/. Accessed June 22, 2024.
16. Smart CE, Ross K, Edge JA, Collins CE, Colyvas K, King BR. Children and adolescents on intensive insulin therapy maintain postprandial glycaemic control without precise carbohydrate counting. *Diabet Med*. 2009;26(3):279-285. doi:10.1111/j.1464-5491.2009.02669.x.
17. MethodCompare: bias and precision plots. https://CRAN.R-project.org/package=MethodCompare. Published 2022. Accessed June 22, 2024.
18. nlme: linear and nonlinear mixed effects models. https://CRAN.R-project.org/package=nlme. Published 2023. Accessed June 22, 2024.
19. emmeans: estimated marginal means, aka least-squares means. https://CRAN.R-project.org/package=emmeans. Published 2023. Accessed June 22, 2024.
20. Myles PS, Cui J. I. Using the Bland–Altman method to measure agreement with repeated measures. *Br J Anaesth*. 2007;99(3):309-311. doi:10.1093/bja/aem214.

21. R: a language and environment for statistical computing. https://www.R-project.org/. Published 2022. Accessed June 22, 2024.
22. Joubert M, Meyer L, Doriot A, Dreves B, Jeandidier N, Reznik Y. Prospective independent evaluation of the carbohydrate counting accuracy of two smartphone applications. *Diabetes Ther*. 2021;12(7):1809-1820. doi:10.1007/s13300-021-01082-2.
23. Rhyner D, Loher H, Dehais J, et al. Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: a comparative study. *J Med Internet Res*. 2016;18(5):e101. doi:10.2196/jmir.5567.
24. Anthimopoulos M, Dehais J, Shevchik S, et al. Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones. *J Diabetes Sci Technol*. 2015;9(3):507-515. doi:10.1177/1932296815580159.
25. Bell KJ, Barclay AW, Petocz P, Colagiuri S, Brand-Miller JC. Efficacy of carbohydrate counting in type 1 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*. 2014;2(2):133-140. doi:10.1016/S2213-8587(13)70144-X.