Published in final edited form as:

Anal Bioanal Chem. 2011 July ; 400(9): 2871–2880. doi:10.1007/s00216-011-5004-5.

Investigation of the specificity of Raman spectroscopy in non-invasive blood glucose measurements

Narahara Chari Dingari, Ishan Barman^{*}, Gajendra P. Singh, Jeon Woong Kang, Ramachandra R. Dasari, and Michael S. Feld[†]

George R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

Abstract

Although several *in vivo* blood glucose measurement studies have been performed by different research groups using near-infrared (NIR) absorption and Raman spectroscopic techniques, prospective prediction has proven to be a challenging problem. An important issue in this case is the demonstration of causality of glucose concentration to the spectral information, especially as the intrinsic glucose signal is smaller compared to that of the other analytes in the blood-tissue matrix. Furthermore, time-dependent physiological processes make the relation between glucose concentration and spectral data more complex. In this article, chance correlations in Raman spectroscopy based calibration model for glucose measurements are investigated for both *in vitro* (physical tissue models) and *in vivo* (animal model and human subject) cases. Different spurious glucose concentration profiles are assigned to the Raman spectra acquired from physical tissue models, where the glucose concentration is intentionally held constant. Analogous concentration profiles, in addition to the true concentration profile, are also assigned to the datasets acquired from an animal model during a glucose clamping study as well as a human subject during an oral glucose tolerance test (OGTT) test. We demonstrate that the spurious concentration profile based calibration models are unable to provide prospective predictions, in contrast to those based on actual concentration profiles, especially for the physical tissue models. We also show that chance correlations incorporated by the calibration models are significantly less in Raman as compared to NIR absorption spectroscopy, even for the *in vivo* studies. Finally, our results suggest that the incorporation of chance correlations for *in vivo* cases can be largely attributed to the uncontrolled physiological sources of variations. Such uncontrolled physiological variations could either be intrinsic to the subject or stem from changes in the measurement conditions.

Keywords

Raman spectroscopy; non-invasive glucose monitoring; chance correlations; causation; animal model; human subject

Introduction

Non invasive measurement of blood and tissue analytes is a very important goal in laboratory medicine and a critical component of point of care diagnostic testing. Specifically, transcutaneous detection of blood glucose will improve the quality of life in diabetic patients [1,2], many of whom must undergo glucose testing several times each day. Such technology will have a major impact on the management of diabetes, including the

^{*}Correspondence to: Ishan Barman, ishan@mit.edu.

[†]Deceased

realization of continuous glucose monitoring systems necessary for the clinical implementation of treatment modalities such as an artificial pancreas [3,4]. In this regard, various optical and spectroscopic methods, such as vibrational spectroscopy (notably NIR absorption and Raman spectroscopy), fluorescence spectroscopy and optical rotations, have been extensively researched [5, 6, 7, 8]. Among these optical methods, the vibrational spectroscopic modalities have demonstrated the greatest promise.

NIR absorption spectroscopy, which probes the overtone and combination vibrational transitions of the molecules, has been reported to provide noninvasive glucose measurements in both healthy and diabetic human subjects [9, 10, 11]. However, such transition bands are typically very broad leading to spectra with a large degree of spectral overlap between the analyte of interest and other interferents. To isolate the concentration information from the analyte of interest in the mixture samples, multivariate calibration (MVC) models such as partial least squares (PLS) regression [12] or support vector regression [13] are applied. These implicit calibration methods are typically used to solve the resultant underdetermined system of equations (where the number of spectral data channels (wavelengths) greatly exceed the number of calibration samples), as they require only calibration spectra from the samples and concentration measurements of analyte of the interest. Unfortunately, the computation of the regression vector by these implicit calibration methods may result in a situation, where functional models are based on chance correlations in the relationship between the spectral information and the glucose concentrations. In particular, Arnold, Xu and co-workers have questioned the validity of NIR absorption spectroscopy based transcutaneous blood glucose measurement reports, especially when the spectra are acquired in a time-dependent fashion [14, 15].

On the other hand, Raman spectroscopy provides excellent molecular specificity as it probes the fundamental vibrational states, which exhibit intrinsically sharper spectral features. Several laboratories including our own have reported promising results for glucose predictions using MVC models in whole blood samples [16], *in vitro* human aqueous humor [17] and even in human volunteers [18, 19]. Despite these preliminary results, successful prospective application in the form of accurate prediction on individual samples not in the calibration set ("prospective prediction") remains elusive. This can be primarily attributed to non-analyte specific variances such as sample absorption and scattering (turbidity) [20], and the resultant incorporation of curved effects in the spectra-concentration relationship [21]. Recently, our laboratory has also addresses other confounding factors include tissue autofluorescence and associated photobleaching [22] and the presence of a physiological glucose lag between the blood and interstitial fluid glucose [23]. Nevertheless, the presence and extent of chance correlations arising from instrumental drift and other unmeasured physiological activities have not been extensively investigated. Clearly, such correlations, in the absence of true causation, can lead to inaccurate predictions, when prospective application is carried out on a different day or on a different individual [24]. Therefore, the importance of the careful validation of the calibration models cannot be overstated.

In this paper, we investigate the impact of chance correlations on Raman spectra based calibration models. We study a set of physical tissue models (tissue phantoms), where the concentration of the analyte of interest (glucose) is intentionally kept constant and the concentrations of other spectral interferents are varied randomly. Similar to the approach described by Arnold and co-workers [25], PLS regression models are constructed by assigning different glucose concentration profiles to the spectral dataset acquired from the tissue phantoms. Here, we have used glucose concentration profiles previously used by the Heise [10] and Arnold [25] groups to enable direct comparison of the MVC models based on Raman spectra to those based on NIR absorption spectra for the same set of assigned concentrations. Our results exhibit that for this *in vitro* case, functional models are not

created based on the spurious glucose concentration profiles, which is in contrast to the results reported in the literature for NIR absorption measurements [25]. We observe that our Raman spectroscopic approach provides surprising robustness to chance correlations particularly those arising from instrumental intensity drifts and measurement condition variations. Further, we employ representative datasets acquired from an animal model in a glucose clamping study and a human subject undergoing an oral glucose tolerance test (OGTT). In these studies, we observe a considerable (adverse) impact of chance correlations showing that glucose-interferent concentration relationship based on uncontrolled physiological variations (either intrinsic to the subject or arising from changes in measurement conditions) cannot be completely disregarded for *in vivo* Raman spectroscopy.

Experimental

In order to investigate the presence of chance correlations in the Raman spectra-based MVC models, three sets of experimental studies are employed: (i) tissue phantom study (ii) animal study and (iii) human subject study. The tissue phantom study provides an ideal platform for the comparison of analyte specificity of Raman spectra to that of NIR absorption spectra in constructing regression models. In addition, a glucose clamping study in an animal model (dog), which features substantially greater complexity in comparison to the well-controlled tissue phantom experiments, is performed to understand the pitfalls that can be encountered in developing glucose-specific regression models in human population using conventional calibration experiments, such as oral glucose tolerance tests (OGTT). Dog models are selected for the current investigation because they provide much greater flexibility in terms of spectral acquisition and clamping study design, as compared to human subject studies, while exhibiting similar physiological glucose response. Finally, a representative dataset from a human subject study [18] is employed here to study the differences introduced by employing a glucose tolerance protocol in contrast to a clamping protocol. Clearly, the latter provides a more randomized run order of collection of spectral data and therefore is less prone to temporal correlations (both instrumental and physiological) than the tolerance test profiles.

Tissue Phantom Study

The preparation and acquisition of the tissue phantom dataset used in this article was detailed in one of our laboratory's previous publications [20]. Briefly, 36 tissue phantoms were prepared in distilled water, where the glucose concentration was kept constant (500 mM) while varying the other analytes (intralipid and India ink). In particular, the intralipid and India ink concentrations were varied to simulate the typical absorption (0.08 to 1.3 cm^{-1}) and scattering coefficients (24 to 130 cm^{-1}) of biological tissue in the tissue phantoms [26]. An 830 nm diode laser was employed as the excitation source and a f/1.4 spectrograph in conjunction with a liquid nitrogen cooled CCD was used to collect Raman spectra from these tissue phantoms. It should be noted that the spectra were acquired randomly with respect to the concentrations of any of the tissue phantom constituents. Indeed, the correlations between the concentrations of any two tissue phantom constituents or between a constituent concentration and the sample run order are observed to be less than 0.1. The acquired spectra are subjected to curvature correction, vertical binning and cosmic ray removal prior to calibration model development detailed in the Data Analysis section. Spectra from 360 – 1600 cm^{-1} are used for all our data analysis (lower wavenumber spectral regions are not considered because of the overlap with the notch filter response and higher wavenumber spectral regions suffer from the deterioration of CCD quantum efficiency).

Animal Model Study

Experiments are conducted in beagle dogs (Fort Wayne, IN) housed under controlled conditions in an animal facility. Here, we employ a representative dataset acquired from one such dog subject undergoing a glucose clamping studies (courtesy of Dr. Mihailo V. Rebec). The dog subject is anesthetized by using isoflurane gas for the period of glucose clamping study (~8 hours). The dog's plasma glucose levels are changed and maintained at various levels ("clamped") by 50% dextrose and insulin infusion through a 22-gauge catheter in the saphenous vein. Another adjacent catheter is used for obtaining blood samples for reference glucose concentration measurements from the same vein. The blood glucose concentrations are measured using an Analox analyzer every 5 minutes during the clamping phase. A rectal thermometer is used to observe the dog's core temperature every 10 minutes. In addition, a blood pressure monitor, pulse oximeter and respiratory monitor are employed to keep track of the vital parameters. All experimental procedures for the dog study are approved by the Purdue Animal Care and Use Committee and the Massachusetts Institute of Technology Committee on Animal Care.

A portable version of our laboratory instrument is used for acquisition of the Raman spectra in this case [27]. The 830 nm excitation laser with approximately 300 mW power is focused from beneath the dog's ear (excitation spot size of 1 mm²) through a hole in the customized paraboloidal mirror [28], which is used to guide the back-scattered light into the spectrograph-CCD combination. The dog's ear is placed on a sapphire window, which serves as a reference plane and, importantly, does not provide a significant Raman or fluorescent background signal. It is worth emphasizing that the optical system employed a free space non-contact configuration and did not employ a fiber probe for excitation and collection of the Raman signal. Raman spectra are collected continuously in this configuration with a total time of 6.5 sec per frame (i.e. acquisition plus data transfer). The collected spectra are subjected to the same pre-processing steps stated previously for the tissue phantom study. Figure 1 gives the spectral time series and the blood glucose concentration time profile measured from this dog subject.

Human Subject Study

Here, we employ a representative dataset from a human subject undergoing an OGTT, which was originally acquired as part of a larger study investigating the pre-clinical feasibility of Raman spectroscopy in the non-invasive detection of blood glucose in healthy human volunteers [18]. In this study, Raman spectra were acquired from the forearms of healthy Caucasian and Asian volunteers undergoing OGTT using a laboratory based bench-top Raman spectrometer. An 830 nm diode laser was used for excitation at an average power of ~300 mW in a ~1mm² spot. All measurements were performed in a non-contact mode from the human volunteers' forearms, which were placed in a specialized mechanical contraption to reduce motional artifacts. Following the initial ingestion of a glucose-rich beverage, Raman spectra were collected every 5 minutes, with an acquisition time of 3 minutes, over a typical 2–3 hour measurement period. The reference blood glucose concentrations were measured using a HemoCue glucose analyzer on finger-pricked blood samples. These human volunteer studies were approved by the Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects.

Data Analysis

Tissue Phantom Study

Three different glucose concentrations are assigned to the aforementioned 36 tissue phantom spectral dataset for construction of the PLS models. The first concentration profile consists of randomly assigned values ranging from 1 to 20 mM, here after referred to as the random

concentration profile. The second and third concentration profiles are based on experimental observations of the Heise [10] and Arnold [25] groups. Figure 2 shows graphically interpolated glucose-time profiles presented in these published reports. It is worth mentioning that while the original figures had 86 and 68 concentration values in their respective profiles, these values have been re-sampled here to provide 36 concentration values consistent with the number of tissue phantoms used in this study. The spectral dataset are not subjected to further processing steps, such as removal of the fluorescence background, in the analysis presented here. The prediction dataset is first created by randomly extracting the data corresponding to 12 tissue phantoms. PLS calibration models are then developed based on the spectra of the remaining 24 tissue phantoms and the respective concentration values. Leave-one-out cross validation is employed within the calibration dataset to optimize the number of loading vectors used in the final PLS regression model. The developed regression model is then used to predict on the 12 samples to compute the standard error of prediction (SEP) in the following manner.

$$SEP = \left[\frac{1}{N} \sum_{i=1}^N (C_a - C_p)^2 \right]^{\frac{1}{2}} \quad (1)$$

where N represents the number of spectra in the prediction dataset, C_a and C_p are the assigned glucose concentrations and the model-predicted glucose concentrations, respectively. In addition, the F -value and regression parameters are computed for comparison of the calibration models constructed for the three different glucose-time profiles. F -test is extensively used to determine whether two normal populations have the same variance and is employed here to assess if the variability of the predicted concentrations is greater than would be expected by chance [29]. For our studies, we compute the following F -test statistic to assess the null hypothesis that the variance of errors of glucose predictions via a PLS model is same as the variance of the assigned glucose concentrations (consistent with the model not explaining any glucose-specific variation):

$$F = \frac{SDP^2}{SEP^2} \quad (2)$$

where SDP is the standard deviation of the assigned concentrations in the prediction dataset. From the F -test statistic and knowledge of number of samples in the prediction set, one can determine the probability level at which the F -value is significant (i.e. characteristic of the confidence level for rejecting the null hypothesis). This probability level is an important metric as it accounts for the number of samples tested and can thus be used to compare between populations with different number of samples, e.g. a F -value of 1.53 calculated from 12 prediction samples is significant at the 76% probability level whereas the same value calculated from 62 samples is significant at a 95% probability level. These parameters also provide significant insight into the relative robustness of the Raman spectra-based calibration models vis-à-vis the NIR absorption spectra-based calibration models, as reported in Arnold *et al.* [25].

Animal Model Study

During the clamping phase, a total of 3330 single-frame spectra are collected and 71 blood glucose concentrations are measured in the animal model study. To ensure compatibility between the two sets of data, consecutive sets of 15 single-frame spectra are averaged, which also enhances the resultant SNR, and the concentration time profile is likewise re-sampled to obtain 222 blood glucose values over the same duration. In addition to the real concentration profile, three spurious concentration profiles are also assigned to the spectral

dataset (akin to the tissue phantom study), namely random, Heise and Arnold. For the latter two, the concentration values are re-sampled from 86 and 68 to 222 concentration values in each case. For all four of these datasets, 20 prediction samples are held aside for validation of the calibration models, which are created from the remainder of the dataset. Similar to the calibration model development for the tissue phantom study, leave-one-out cross-validation is used to tune the number of loading vectors by minimizing the cross-validation error. Using the final regression models on the prediction samples, the SEP, F -value and the linear regression parameters are calculated.

Human Subject Study

In the representative human subject dataset, 25 blood glucose concentrations and corresponding Raman spectra were acquired. Similar to the animal model study, two spurious concentration profiles are assigned to the human subject spectral dataset, namely a random profile and that of Arnold. Unlike the tissue phantom and animal model study we do not assign the Heise glucose concentration profile to our spectral dataset here, because it substantially resembles our measured concentration profile during the tolerance test. This is not surprising because both the experiments were performed using the standard glucose tolerance protocol. The high correlation between the profiles ($r = 0.90$) hampers the suitability of the Heise concentration profile in probing chance correlations for spurious glucose profiles that do not have any similarity with the actual one. The concentration values for the Arnold profile is re-sampled from 68 points to 25 points. Due to the lack of sufficient number of samples in this case, leave-one-out cross-validation (LOOCV) in place of prospective prediction is used to assess the performance of the calibration models.

Results and Discussion

Tissue Phantom Study

As mentioned in the data analysis section, three sets of calibration models are developed on spectra acquired from the tissue phantom study. For all cases, we ensured that the number of calibration samples used (24) is at least three times larger than the model rank, namely the number of loading vectors incorporated in the model [30]. It was ensured during the splitting of the entire dataset into calibration and prediction subsets that the calibration dataset spanned a sufficiently wide range of tissue phantom parameters to allow the creation of a model capable of prospective prediction. Specifically, we note that the absorption and reduced scattering coefficients for the calibration dataset had the same ranges of 0.08–1.3 cm^{-1} and 3.68–17.48 cm^{-1} as the prediction dataset. Additionally, the mean and standard deviation of the absorption and reduced scattering coefficients of the calibration dataset are observed to be within 15% of the corresponding values for the prediction dataset, providing further evidence that the calibration data forms an adequate representation of the randomly selected prediction samples. (It should be noted that the concentration of the analyte of interest, glucose, was held constant in all the tissue phantoms.)

Figure 3 plots the glucose prediction results in the tissue phantom datasets - using the random (red circle), Heise (blue circle) and Arnold (magenta circle) concentration profiles - on a Clarke error grid [31], which is widely used to quantify the clinical accuracy of blood glucose measurements. The grid plots predicted and reference glucose concentrations in five different regions, where regions A and B are considered acceptable for determining therapeutic options, whereas regions C, D and E are not reliable and may be risky for clinical decisions. Since concentration of glucose in all the tissue phantoms is kept constant, one would ideally expect that the prediction would not correlate with the spurious concentration profiles assigned in PLS modeling. Clearly, the PLS model developed from the random glucose concentration-time profile is unable to estimate glucose concentrations

in the prediction dataset. Specifically, we observe that the SDP is approximately 6.63 mM, where as the standard error of prediction (SEP) is 6.7 mM. Given the variances of these two distributions, the F -value is found to be significant only at the 52% probability level, i.e. the null hypothesis that the two distributions have the same variance cannot be rejected. In other words, the calibration model predictions have a fairly weak correlation ($r = 0.34$) with the reference values, which is promising given that there is no such variance in the experimental model. The linear regression parameters stated in Table 1 further validates this understanding.

Similarly, Fig. 3 also suggests that the PLS models developed from the non-random Heise and Arnold concentration profiles display an inability to predict the reference glucose concentrations. The SDP and SEP values for Heise concentration profile is 7.47 mM and 6.37 mM respectively and the F -value is significant at the 70% probability level. In the same way, SDP and SEP values for Arnold concentration profile are 4.63 mM and 4.74 mM, resulting in a F -value which is significant at the 53% level. In addition, the r values of Heise and Arnold concentration profiles are 0.61 and 0.38 respectively, which shows below average-poor correlation of the predicted concentrations with the reference glucose values. Other performance metrics such as linear regression coefficients are listed in Table 1. For all of the concentration profiles, the y-intercept (β_0) is closer to the mean value of the glucose concentration in the prediction dataset than to zero, and the corresponding slopes (β_1) are considerably lower than unity. In summary, we observe that the aforementioned calibration models do not provide 'accurate' glucose predictions based on non-analyte specific variances, thereby indicating that they are robust to chance correlations in the studied tissue phantoms.

To put our results in better perspective, we compare the performance of our tissue phantom calibration models with similar studies reported in the literature for NIR absorption spectroscopy. In particular, using NIR absorption spectra acquired from similar tissue phantom studies, the F -values were calculated to be significant at the 60%, 100% and 89% levels for the random, Heise and Arnold concentration profiles, respectively [25]. Evidently, for the non-random Heise and Arnold profiles, the F -value significance level is substantially higher for the NIR absorption spectra than our tissue phantom Raman spectral dataset. This suggests that Raman spectra are more robust to chance correlations than the corresponding NIR absorption spectra, especially when applied to glucose measurements in similar tissue models. Although the number of prediction samples is different in the published report from our case, it is notable that there is a concomitant increase in correlation coefficient between the predicted and the reference glucose concentration for the absorption spectra based calibration models using the Heise (reported $r = 0.91$) and Arnold profiles (reported $r = 0.73$). These results can be primarily attributed to the fact that Raman spectra have intrinsically sharper bands allowing a better fingerprinting approach in comparison to the NIR absorption spectra, notwithstanding the higher signal strength of the latter.

Importantly, this demonstrates that there is negligible possibility of constructing MVC models based on time-dependent spectral features arising from instrumental intensity drifts and reasonable measurement condition variations, such as room temperature and humidity. We note that larger changes in measurement conditions, especially ambient temperature, may provide significant spectral distortions (in the form of change in peak intensity, peak broadening or/and shift in peak frequency) [32] that can introduce curved effects in the spectra-concentration relationship. This gives rise to a completely separate class of calibration problems, which must be handled by nonlinear MVC approaches such as support vector machines [21]. Additionally, since the spectral interferences are observed to have little or no correlation with one another, with the experiment run order or the assigned glucose concentration profiles ($r < 0.1$ for all possible combinations), we can ignore the possibility of

chance correlations arising from such sources, i.e. spectral interferents and experiment run order. While the reproducibility of the instrumentation hardware ensures that system drift and measurement condition variations (within reasonable ranges) should not significantly affect the *in vivo* studies as well, the dynamic and uncontrolled changes of the interferents inherent in the animal model and human subject studies need to be carefully examined.

It is worth mentioning that the tissue phantoms employed by the Arnold group potentially could have greater complexity than ours, which may have introduced other chance correlations for the same non-randomized time profiles used in our study. The resulting spectral datasets are different in the two studies precluding a direct comparison of the model robustness in the two cases. Nevertheless, it is reasonable to suggest that there is a consistent pattern of behavior showing the superior robustness of Raman spectra-based models to non-random concentration-time profiles. To investigate the issue of robustness versus chance correlations further, we employ the animal model study, where the complexity of the system is greatly enhanced by the inherent physiological and temporal variations.

Animal Model Study

In this study, four calibration models are generated corresponding to the random, Heise, Arnold and real clamping concentration profiles. Figure 4 presents plots for the predicted versus reference concentrations for the four calibration models on Clarke error grids. Expectedly, when the random glucose profile is assigned, we observe that the glucose predictions are scattered randomly on the Clarke error grid showing that the calibration model does not explain glucose information. SDP and SEP values for the random concentration profile is observed to be 7.06 mM and 7.11 mM, which results in a *F*-value significant at 51% probability level and a correlation coefficient *r* of 0.24 (Table 2).

Similarly, the performance parameters for Heise (SDP = 6.08 mM, SEP = 4.88 mM and *F*-value significance at 83%) and Arnold (SDP = 2.96 mM, SEP = 2.36 mM and *F*-value significance at 84%) concentration profiles are evaluated. It is evident that while the calibration models based on the Heise and Arnold profiles still do not fully explain the glucose information, there is a substantial rise in the (adverse) impact of chance correlations, as compared to our tissue phantom study. The decrease in robustness from the tissue phantom to the animal model study can be attributed to the presence of unmeasured physiological variations in the blood tissue matrix of the dog subject. Nevertheless, the above calibration models for the animal model still provide comparable (Arnold) or better (Heise) robustness when viewed in light of the NIR absorption calibration models of the tissue phantom studies mentioned above.

In general, the aforementioned uncontrolled physiological variations could either be intrinsic to the subject or stem from changes in the measurement conditions, including body posture and pressure at sample interface. While changes in measurement conditions are less likely to lead to significant physiological variations in a controlled clamping study environment, one must account for this source of error in more general clinical settings. Another potential source of error in these *in vivo* studies as well as the human subject studies discussed below is the presence of inconsistency between the acquired spectrum (which is primarily indicative of the interstitial fluid glucose level in the dog's ear) and the reference concentration measurement (which represent the systemic blood glucose level). This inconsistency has two primary components, namely the presence of physiological lag between blood and interstitial fluid glucose at any tissue site and the variation in glucose levels from site to site. However, the presence of a clamping protocol in the animal model study substantially reduces such inconsistencies as the clamps allow the glucose levels to equilibrate. A validation check using microporation on the dog's ear revealed that, except when the rate of change in glucose concentration is very high (>3 mg/dL/min), the ISF

glucose in the dog's ear has a very strong positive correlation with the systemic blood glucose measurements ($r > 0.95$). (For the microporation method, a combination of pressure and laser-based thermal ablation was employed to disrupt the stratum corneum of the skin (creating a small hole of 200 μm diameter) in order to obtain ISF from the dermis.)

In contrast to the fictitious profiles, when the real glucose concentration profile (measured during the clamping phase) is assigned, the calibration model exhibits excellent performance as visualized in the Clarke error grid (Fig. 4D). The corresponding SDP and SEP values are noted to be 4.16 mM and 1.63 mM, resulting in a F -value significant at 99.99% probability level. In other words, the calibration model is able to predict the glucose concentrations of the real clamping profile with high accuracy ($r = 0.94$). The substantial disparity in the computed F and r values between the real case and that of the spurious concentration profiles is promising for non-invasive glucose detection studies using Raman spectroscopy.

Human Subject Study

In this study, calibration models are generated corresponding to the random, Arnold and OGTT concentration profiles and used to predict the glucose concentrations in a leave-one-out manner. Table 3 shows the performance parameters for the various calibration models corresponding to the three assigned concentration profiles. Similar to our observations in the animal model study, we see that the random and actual concentration profiles provide the worst and best performances, respectively. Specifically, the SDP (8.01 mM) and SEP (7.62 mM) values for the former do not differ significantly, whereas the SEP (1.06 mM) for the latter shows statistically significant differences (F -value = 99.55 %) with respect to the SDP (1.82 mM) value.

The non-randomized Arnold concentration profile based study apparently provides functional models which explain the glucose information as demonstrated by the computed F -value which is significant at the 94% level. One can readily infer the presence of chance correlations in such a model given that the Arnold concentration profile is not an accurate representation of the actual blood glucose concentration of the human subject. It is notable that the human subject study features a higher F -value for the Arnold profile in comparison to the dog glucose clamping study, which can be attributed to the higher possibility of temporal correlations in a glucose tolerance profile. Nevertheless, the actual concentration profile based calibration models provide a considerably better performance than any of the spurious assigned profiles as seen by the respective correlation coefficients (Table 3).

Finally, as noted by several investigators, other time-dependent factors (i.e. not arising from a glucose interferent concentration relationship) can also play a part in strongly correlating with specific concentration profiles, e.g. from diabetic patients undergoing glucose tolerance testing. We have previously investigated such a phenomenon, especially with respect to Raman spectroscopy, as apparently functional models can be constructed due to the correlation between the glucose concentrations and the fluorescence levels in the acquired spectra that photobleach over time [21]. In this study, the non-monotonic nature of the profiles substantially reduces the possibility of correlating with the quenched fluorescence levels.

In summary, we conclude that Raman spectra-based calibration models appear to have significant robustness, particularly when compared with the similar models constructed from NIR absorption spectra. However, we also observe that increasing the complexity of the system, primarily arising from uncontrolled physiological variations, introduces some chance correlations, which induces false predictions. To overcome this problem, we recommend that at least a double glucose tolerance test approach, where an additional glucose bolus is provided after the first insulin mediated decrease in glucose levels, should

be used for calibration studies in human population since it greatly reduces the possibility of time-dependent correlations. For our tissue phantom study (where the glucose concentration is unchanged in all the samples), assigning a double OGTT profile provides nearly identical results to a random concentration profile - thereby characterizing the effectiveness of a double OGTT profile in precluding the possibility of incorporation of chance correlations in the model. This is further validated by our observation that the correlation coefficient between the predicted and the assigned concentrations in the tissue phantom study typically falls by a factor of two or more when a double OGTT profile is assigned in place of a single OGTT profile, showing the greater robustness of a double OGTT profile. In addition, incorporation of calibration data from several days on the same subject is likely to reduce such factors as one would not expect a consistent pattern of glucose-interferent concentration relationship across multiple days [15].

Conclusions

In this article, the specificity of Raman spectroscopy has been investigated for *in vitro* as well as *in vivo* glucose measurements. Our Raman spectroscopy-based results clearly demonstrate that spurious assigned glucose concentration profiles are unable to predict the real concentration profiles with significant accuracy, especially for *in vitro* models. We also report that Raman spectroscopy is less prone to chance correlations incorporated by the calibration model compared to NIR absorption spectroscopy. Nevertheless, the presence of such correlations based on uncontrolled physiological variations under *in vivo* conditions needs to be carefully examined and accounted for by appropriate study design.

Acknowledgments

This work was supported by the NIH National Center for Research Resources (Grant No. P41-RR02594) and a grant from Bayer HealthCare, LLC. The animal model study was performed at the Indiana University-Purdue University Fort Wayne facility in collaboration with the Bayer HealthCare, Diabetes Care division. Specifically, the animal model dataset used in this article was acquired by Dr. Mihailo V. Rebec and his clinical team. One of the authors, IB, acknowledges the support of Lester Wolfe Fellowship from the Laser Biomedical Research Center.

References

1. Klonoff DC. Diabetes Care. 1997; 20:433–437. [PubMed: 9051401]
2. Roe JN, Smoller BR. Crit Rev Ther Drug Carrier Syst. 1998; 15:199–241. [PubMed: 9699080]
3. Sullivan SJ, Maki T, Borland KM, Mahoney MD, Solomon BA, Muller TE, Monaco AP, Chick WL. Science. 1991; 252:718–721. [PubMed: 2024124]
4. Charles MA. Diabetes Technology & Therapeutics. 1999; 1:89–96. [PubMed: 11475311]
5. Khalil OS. Diabetes Technology & Therapeutics. 2004; 6:660–697. [PubMed: 15628820]
6. Tuchin, VV. Handbook of Optical Sensing of Glucose in Biological Fluids and Tissues. CRC Press; 2009.
7. Berger AJ, Itzkan I, Feld MS. Spectrochim Acta A. 1997; 53:287–292.
8. Cote GL, Fox MD, Northrop RB. IEEE Trans Biomed Eng. 1992; 39:752–756. [PubMed: 1516942]
9. Maruo K, Tsurugi M, Tamura M, Ozaki Y. Appl Spectrosc. 2003; 57:1236–1244. [PubMed: 14639751]
10. Heise HM, Marbach R, Koschinsky TH, Gries FA. Artif Organs. 1994; 18:439–447. [PubMed: 8060253]
11. Samann A, Fischbacher CH, Jagemann KU, Danzer K, Schuler J, Papenkordt L, Muller UA. Exp Clin Endocrinol Diabetes. 2000; 108:406–413. [PubMed: 11026754]
12. Brereton, RG. Applied Chemometrics for Scientists. John Wiley & Sons Ltd; Chichester, West Sussex, England: 2007.
13. Thissen U, Ustun B, Melssen WJ, Buydens LMC. Anal Chem. 2004; 76:3099–3105. [PubMed: 15167788]

14. Arnold MA. *Curr Opin Biotech.* 1996; 7:46–49. [PubMed: 8742375]
15. Liu R, Chen W, Gu X, Wang RK, Xu K. *J Phys D: Appl Phys.* 2005; 38:2675–2681.
16. Berger AJ, Koo TW, Itzkan I, Horowitz G, Feld MS. *Appl Opt.* 1999; 38:2916–2926. [PubMed: 18319874]
17. Lambert JL, Pelletier CC, Borchert M. *J Biomed Opt.* 2005; 10:031110–031118. [PubMed: 16229635]
18. Enejder AMK, Scecina TG, Oh J, Hunter M, Shih WC, Sasic S, Horowitz GL, Feld MS. *J Biomed Opt.* 2005; 10:031114-1-031114-9. [PubMed: 16229639]
19. Chaiken J, Finney W, Knudson PE, Weinstock RS, Khan M, Bussjager RJ, Hagrman D, Hagrman P, Zhao Y, Peterson CM, Peterson K. *J Biomed Opt.* 2005; 10:031111-1–031111–12. [PubMed: 16229636]
20. Barman I, Singh GP, Dasari RR, Feld MS. *Anal Chem.* 2009; 81:4233–4240. [PubMed: 19413337]
21. Barman I, Kong CR, Dingari NC, Dasari RR, Feld MS. *Anal Chem.* 2010; 82:9719–9726. [PubMed: 21050004]
22. Barman I, Kong CR, Singh GP, Dasari RR. *J Biomed Opt.* 2010; 16:011004-1-011004-10. [PubMed: 21280891]
23. Barman I, Kong CR, Singh GP, Dasari RR, Feld MS. *Anal Chem.* 2010; 82:6104–6114. [PubMed: 20575513]
24. Liu R, Deng B, Chen W, Xu K. *Opt Quant Electron.* 2005; 37:1305–1317.
25. Arnold MA, Burmeister JJ, Small GW. *Anal Chem.* 1998; 70:1773–1781. [PubMed: 9599578]
26. Cheong W, Prah S, Welch AJ. *IEEE J Quantum Electron.* 1990; 26:19.
27. Shih, WC. Quantitative biological Raman spectroscopy for non-invasive blood analysis. Massachusetts Institute of Technology, Dept. of Mechanical Engineering; 2007.
28. Enejder AMK, Koo TW, Oh J, Hunter M, Sasic S, Feld MS, Horowitz GL. *Opt Lett.* 2002; 27:2004–2006. [PubMed: 18033426]
29. Johnson, NL.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions.* Vol. 2. Wiley; 1996.
30. Qi D, Berger AJ. *Appl Opt.* 2007; 46:1726–1734. [PubMed: 17356615]
31. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. *Diabetes Care.* 1987; 10:622–628. [PubMed: 3677983]
32. Wulfert F, Kok WT, Smilde AK. *Anal Chem.* 1998; 70:1761–1767. [PubMed: 21651271]

Figure 1. Representative Raman spectra acquired from a dog model during a glucose clamping study and corresponding blood glucose concentrations measured over the same time.

Figure 2. Glucose concentration-time profiles used for our tissue phantom and dog model study. The data here is re-drawn from published reports by the Heise (left) [10] [with permission from John Wiley and Sons] and Arnold [25] (right) [with permission from American Chemical Society] groups.

Figure 3. Glucose predictions for the three different calibration models developed for the tissue phantom study using the random (red circle), Heise (blue circle) and Arnold (magenta circle) concentration profiles, respectively, plotted on a single Clarke Error Grid.

Figure 4.

Glucose predictions for the four different calibration models developed for the dog subject study using the random (A), Heise (B), Arnold (C) and real (D) concentration profiles plotted on the Clarke Error Grid.

Table 1

Performance characteristics for tissue phantom calibration models

Profile	Regression analysis of the prediction data					
	SDP (mM)	SEP (mM)	Model size	F- prob. (%)	β_0 (mM)	β_1 r
Random	6.63	6.71	12	52	8.57	0.28 0.34
Heise	7.47	6.37	12	70	5.82	0.61 0.61
Arnold	4.63	4.74	12	53	8.70	0.35 0.38

Table 2

Performance characteristics for animal model calibration model

Profile	-				<u>Regression analysis of the prediction data</u>		
	SDP (mM)	SEP (mM)	Model size	F-prob. (%)	β_0 (mM)	β_1	r
Random	7.06	7.11	20	51	11.80	0.10	0.24
Heise	6.08	4.88	20	83	4.43	0.51	0.68
Arnold	2.96	2.36	20	84	4.79	0.62	0.67
Real (Dog)	4.16	1.63	20	99.99	1.33	0.94	0.94

Table 3

Performance characteristics for human subject study

Profile	Regression analysis of the prediction data					
	SDP (mM)	SEP (mM)	Model size	F- prob. (%)	β_0 (mM)	β_1 r
Random	8.01	7.62	25	60	11.30	0.15 0.31
Arnold	4.82	3.53	25	94	4.23	0.66 0.70
Human	1.82	1.06	25	99.55	2.85	0.67 0.81