Statistical Frailty Modeling for Quantitative Analysis of Exocytotic Events Recorded by Live Cell Imaging: Rapid Release of Insulin-Containing Granules is Impaired in Human Diabetic β-cells

Giuliana Cortese[1], Nikhil Gandasi[2], Sebastian Barg[2], Morten Gram Pedersen[3*]

1 Department of Statistical Sciences, University of Padua, Padua, Italy
2 Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden
3 Department of Information Engineering, University of Padua, Padua, Italy

* Corresponding author. E-mail: pedersen@dei.unipd.it

Abstract

Hormones and neurotransmitters are released when secretory granules or synaptic vesicles fuse with the cell membrane, a process denoted exocytosis. Modern imaging techniques, in particular total internal reflection fluorescence (TIRF) microscopy, allow the investigator to monitor secretory granules at the plasma membrane before and when they undergo exocytosis. However, rigorous statistical approaches for temporal analysis of such exocytosis data are still lacking. We propose here that statistical methods from time-to-event (also known as survival) analysis are well suited for the problem. These methods are typically used in clinical settings when individuals are followed over time to the occurrence of an event such as death, remission or conception. We model the rate of exocytosis in response to pulses of stimuli in insulin-secreting pancreatic β-cell from healthy and diabetic human donors using piecewise-constant hazard modeling. To study heterogeneity in the granule population we exploit frailty modeling, which describe unobserved differences in the propensity to exocytosis. In particular, we insert a discrete frailty in our statistical model to account for the higher rate of exocytosis in an immediately releasable pool (IRP) of insulin-containing granules. Estimates of parameters are obtained from maximum-likelihood methods. Since granules within the same cell are correlated, i.e., the data are clustered, a modified likelihood function is used for log-likelihood ratio tests in order to perform valid inference. Our approach allows us for example to estimate the size of the IRP in the cells, and we find that the IRP is deficient in diabetic cells. This novel application of time-to-event analysis and frailty modeling should be useful also for the study of other well-defined temporal events at the cellular level.

Introduction

Novel methods for the study of cell biological processes produce unprecedented data to be analyzed. To maximize the information that can be extracted from the experimental results, appropriate and advanced statistical analytical methods should be exploited. Recent microscopy techniques, in particular total internal reflection fluorescence (TIRF) microscopy, have made it possible to visualize single exocytotic events in neurons and endocrine cells [1–8]. Exocytosis is the process during which the lipid membranes of neurotransmitter-filled synaptic vesicles (in neurons), or hormone-containing secretory

granules (in endocrine cells), fuse with the cell membrane, which allows the signalling molecules contained within the granule to escape to the extracellular space [9].

Such imaging data has given deep insight into the molecular and dynamical regulation of exocytosis. However, to our knowledge, these single-granule data have until now been analyzed by counting the (cumulative) number of events over all observed granules and cells, sometimes followed by simple curve fitting [3, 4, 10]. Thus, more rigorous methods for quantification and analysis of imaging data of exocytosis are needed [6]. We propose here that the detailed temporal information contained in this type of data allows for statistical analysis using tools from time-to-event (also known as survival or failure time) analysis. These methods are typically used for clinical or demographic data where individuals are followed until a certain event of interest, such as death, onset of disease, conception, first-time marijuana use, etc. [11–13]. Another common area of their application is reliability engineering where the interest is the time to failure of an instrument or machine. Since the structure in such data is similar to the live cell imaging exocytosis data of interest here, it allows us to apply these well-established statistical methods on completely different biological and temporal scales.

Insulin is released from pancreatic β-cells in response to various stimuli, with glucose being the physiologically most important. Disturbed insulin secretion is now recognized as a central player in the development of diabetes, a devastating disease which is reaching epidemic proportions [14, 15]. Glucose is transported into the β-cells where it triggers a complex cascade of events leading to cell depolarization and electrical activity. As a result, voltage-dependent Ca^{2+} channels open, promoting Ca^{2+} influx, and the increase in intracellular Ca^{2+} levels cause exocytosis of insulin-containing secretory granules [16]. Insulin secretion is biphasic in response to a sustained glucose stimulus; a large peak of insulin release is followed by a second phase where insulin is released in distinct pulses [17]. Importantly, biphasic insulin release is disturbed in diabetes [18], which has been suggested to have its origin within the pancreatic β-cells [19], likely because of dysfunctional exocytosis [20–22].

It was early proposed that heterogeneous release propensities of the insulin-containing granules could underlie the biphasic secretion pattern. In this hypothesis a small pool of granules is released to yield the first peak of insulin whereas slower release of other granules produce the second phase of secretion [23]. More recent results in various endocrine cells [24–26] showed that a sustained elevation of intracellular Ca^{2+} levels could produce a phasic exocytosis pattern as measured by membrane capacitance recordings reflecting whole-cell release. Also, repeated or sustained depolarizations, which promote Ca^{2+} entry via voltage-dependent Ca^{2+} channels, triggered phasic capacitance patterns, even when investigated as a function of Ca^{2+} entry [16, 27, 28]. These patterns were interpreted as the results of depletion of a small immediately releasable pool (IRP) of granules followed by slower release from a larger pool. Various mathematical models of granule pools and exocytosis were developed based on these and similar results, with the scope of reproducing and simulating typical behavior, in order to investigate the underlying biological mechanisms [25, 29–33]. However, the aim of such mathematical models is not to extract information from raw experimental data. For such a task, statistical methods are needed.

We propose and show here that survival analysis methods can be advantageously applied to cell biological data to provide statistically sound results on completely different biological and temporal scales than their typical areas of application. In particular, we apply time-to-event analysis to exocytosis data from healthy and diabetic human β-cells to quantify hazards (rates of exocytosis) and heterogeneity. In survival analysis, univariate frailty modeling is a method to take into account unobserved differences in hazards between individuals [13, 34]. In the present context, imaging of

the secretory granules can not reveal their release propensity, i.e., whether they belong to the IRP. Based on the biological findings and interpretations cited above, we thus allow for heterogeneity by including frailties in our statistical model. This approach allows us to estimate the size of the IRP directly from single-granule exocytosis data. We estimate that the IRP is smaller in diabetic cells, and that exocytosis is less tightly controlled by depolarizing K^+ pulses compared to healthy cells.

Materials and Methods

Data description

Human pancreatic islets were provided by the Nordic Network for Clinical Islet Transplantation (Uppsala, Sweden) with full ethical approval (Regionala Etikprövningsnämden, Uppsala). Islets were dissociated into single cells in 0.0025% trypsin in Ca^{2+}/Mg^{2+}-free cell dissociation buffer (ThermoFisher) for 3-5 minutes and seeded onto polylysine-coated glass coverslips, and cultured in CMRL 1066 medium containing 5.6 mM glucose, 10% fetal calf serum (FCS), and 2 mM L-glutamine, streptomycin (100 μg/ml), penicillin (100 μg/ml). Seeded cells were infected using adenovirus encoding the granule marker NPY-mCherry (Neuropeptide Y fused to the red fluorescent protein mCherry; [5, 35]) and imaged 24-36 hours later. Insulin-containing secretory granules in pancreatic β-cells from 3 healthy (11 cells) and 2 diabetic (8 cells) donors were imaged using total internal reflection fluorescence (TIRF) microscopy at a frame rate of 10 Hz, with excitation at 561 nm and emission at 590-630 nm. Cells were bathed in (in mM) 138 NaCl, 5.6 KCl, 1.2 MgCl2, 2.6 CaCl2, 10 D-glucose and 5 HEPES (pH 7.4 with NaOH), 2 μM forskolin and 200 μM diazoxide. The latter prevents glucose-dependent depolarization by opening ATP-dependent K^+-channels. Forskolin, which increases intracellular cyclic-AMP, was routinely included to increase the number of primed granules available for exocytosis.

Exocytosis was then evoked by ten 1-second long pulses of local application of high concentrations of K^+ (75 mM KCl equimolarly replacing NaCl), interspersed by 9-seconds long rest intervals (Fig 1). The K^+ pulses depolarize the cellular membrane potential within $\sim$50 ms (unpublished observation), which opens voltage-dependent Ca^{2+} channels and the resulting Ca^{2+} influx triggers exocytosis. The rate of exocytosis is therefore expected to be higher during, compared to between, K^+ pulses. All experiments were carried out with constant buffer perifusion at 32 °C. Exocytosis events were found manually as sudden disappearance of labeled granules.

We considered the granules within a cell as a cluster of statistical units indexed by $j = 1, \ldots, J$. Our data contains $J = 19$ clusters corresponding to the 19 cells, i.e., the terminology 'cluster' refers to a structure in the data: the observations obtained from the granules (the statistical units) in a cell. Cluster j had n_j observations, representing the granules in the cell, with index $i = 1, \ldots, n_j$. For granule i in cell j we observed either the time of exocytosis, $\tilde{t}_{ij}$, or the *censoring time* c_{ij}, i.e., the last observed time. In these data, c_{ij} is the time when the experiment ended, and is thus the same for all granules (so-called administrative censoring). Censoring precludes the observation of exocytosis that might have occurred at a later time. Thus, the observed data are the pairs (t_{ij}, d_{ij}), where t_{ij} are the realizations of the observed survival time $T_{ij} = \min(\tilde{T}_{ij}, C_{ij})$, and d_{ij} is the observed indicator from $D_{ij} = I(\tilde{T}_{ij} < C_{ij})$ that tells whether a granule underwent exocytosis ($d_{ij} = 1$) or was censored ($d_{ij} = 0$). This form of the data is typical for time-to-event data.

Poisson regression modelling

For the analysis of the exocytosis data, we proceeded progressively. Poisson regression neglecting heterogeneity was exploited to investigate whether the data can be described with a time-varying, piecewise constant hazard, although biologically unlikely as discussed below. This approach also serves as the basis for the formulation of the frailty model in the next subsection, as well as a reference frame for the results that follow.

We assumed that the rate (or *hazard function*) of exocytosis $\mu(t)$ is piecewise constant. The hazard was assumed to be constant during each pulse and during each interval between two consecutive pulses, but it may vary from one pulse to another, and from pulses to intervals between pulses. Model selection led to three parameters, ρ_0, ρ_1, ρ_2, estimating, respectively, the rate during the first pulse, the rate during the other pulses, and the rate between pulses. The model also included a covariate X indicating whether the cell came from a healthy ($X = 0$) or diabetic donor ($X = 1$). The effect of diabetes was assumed to be time-varying in a piecewise-constant fashion corresponding to the hazard, i.e., we considered three parameters $\beta_0, \beta_1, \beta_2$ describing the effects, respectively, during the first pulse, during the following pulses, and between pulses. The hazard function was thus modeled as

$$\mu(t|X) = \begin{cases} \rho_0 e^{\beta_0 X} = e^{\alpha_0 + \beta_0 X}, & 0 \leq t < 1 \text{ seconds,} \\ \rho_1 e^{\beta_1 X} = e^{\alpha_1 + \beta_1 X}, & s \leq t < s + 1 \text{ for some } s = 10, 20, 30, \ldots \text{ seconds,} \\ \rho_2 e^{\beta_2 X} = e^{\alpha_2 + \beta_2 X}, & s + 1 \leq t < s + 10 \text{ for some } s = 0, 10, 20, \ldots \text{ seconds,} \end{cases}$$

where $\alpha_k = \log \rho_k$, and $k = 0, 1, 2$ indicate whether t falls in the first pulse ($k = 0$), in one of the following pulses ($k = 1$), or between pulses ($k = 2$) (Fig 1). In particular, we were interested in the question of whether the rate of exocytosis was different between healthy and diabetic cells, and if this difference was restricted to the first pulse.

Since only a small fraction of granules exhibited exocytosis during the experiments, Poisson modeling can be used to describe the data [36]. We used the R [37] function `glm` to perform the analysis. To get cluster-corrected standard errors and Wald-type confidence intervals (which are calculated from standard errors) for the parameter estimates, we used the robust sandwich estimator (see Eq. 5 below) based on R code by Arai [38]. Cox proportional hazards modeling can also investigate the time-dependent effect of diabetes by including time-varying parameters [12], but the baseline hazard function is estimated nonparametrically. When we applied this model, it gave virtually identical results to the Poisson model for the diabetes effect.

Frailty modelling of two pools of granules

The interpretation of the selected Poisson model is that for any granule the rate of exocytosis is higher during the first pulse than during the following pulses, for example because of a reduction in the triggering Ca^{2+} signal as a result of Ca^{2+} channel inactivation. Such an interpretation is biologically unlikely, since the 9 sec interval between pulses is sufficiently long to allow reactivation of Ca^{2+} currents [39]. Thus, if anything, the Ca^{2+} levels should build up from one K^+ pulse to the next, which would increase the rate of exocytosis for pulses later in the train.

An alternative and widely used explanation is to attribute the greater amount of release in the beginning of the stimulus protocol to an immediately releasable pool (IRP) of granules that have a much higher intrinsic rate of exocytosis than the remaining, non-IRP, granules [21, 23]. Once this pool is empty, exocytosis proceeds at a slower pace.

Imaging of the labeled granules can not reveal whether a given granule belongs to the IRP, nor can the size of the IRP be seen from the microscopy images. Statistically,
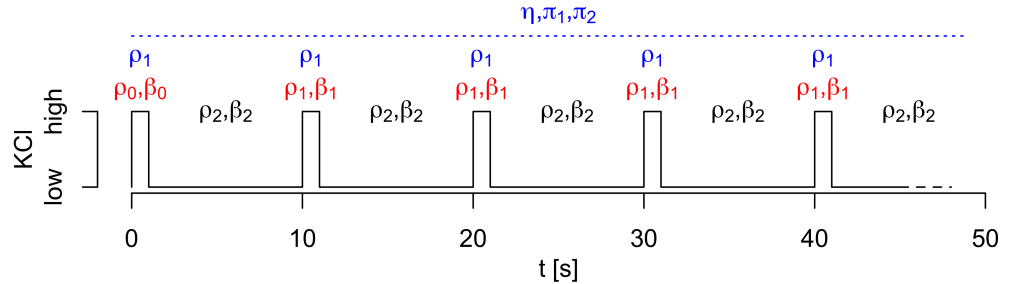
Fig 1. Stimulation protocol and related model parameters. An indication of
the high-concentration K^+ pulses (1 s) interspersed by 9 seconds of rest. The
parameters common to the two statistical models are indicated in black (for the pulses
following the first pulse) and gray (for the rest intervals). The two models have different
parameters for the first pulse. In the Poisson model (red), the baseline rate and effect of
diabetes is allowed to be different during the first pulse compared to subseequent pulses
(black). In the frailty model (blue), the baseline parameters are the same during all the
pulses, but additional parameters (η, π_1, π_2) describing the frailty distribution are
included. These additional parameters are not restricted to a certain time interval. See
main text for detailed descriptions of the statistical models.

we can handle this scenario by introducing a (non-observable) Bernoulli variable Y,
where the realization Y_{ij} is equal to 1 when granule i of cell j belongs to the IRP and 0
otherwise. To allow for different sizes of the IRP in healthy and diabetic cells we
assume that the probability $P(Y = 1|X) = \pi_X$ depends on the diabetes-covariate X.

Exocytosis of an IRP granule is assumed to occur with a rate that is η times higher
than the baseline rate describing non-IRP exocytosis. This assumption is described by a
discrete frailty Z, which takes the value η when $Y = 1$, and $Z = 1$ otherwise. The
resulting frailty model is thus

$$\mu(t|X, Z) = Z\mu_0(t|X), \qquad P(Z = \eta|X) = \pi_X, \quad P(Z = 1|X) = 1 - \pi_X. \quad (1)$$

The baseline hazard μ_0 is piecewise constant with rate $\rho_1 e^{\beta_1 X} = e^{\alpha_1 + \beta_1 X}$ during K^+
pulses and rate $\rho_2 e^{\beta_2 X} = e^{\alpha_2 + \beta_2 X}$ between pulses. Thus, β_1 and β_2 describe effects of
diabetes on the rates-of-exocytosis during and between K^+ pulses, respectively. Note
that in contrast to the Poisson model, the baseline rate is assumed to be identical
during the first and the subsequent K^+ pulses (Fig 1).

In time-to-event analysis, one of the main overall summary measures of interest is
the survival probability $S(t) = P(T \geq t)$, or, equivalently, the cumulative incidence
probability defined as $F(t) = P(T < t) = 1 - S(t)$. S can be estimated in a model-free,
nonparametric way using for example the Kaplan-Meier estimator [11]. For the frailty
model (1), the marginal survival function is given as

$$S(t|X) = \pi_X e^{-\eta M_0(t|X)} + (1 - \pi_X)e^{-M_0(t|X)},$$

where $M_0(t|X) = \int_0^t \mu_0(t|X)$ is the cumulative baseline hazard [13]. This expression is a
mixture of the survival functions of an IRP granule and a non-IRP granule, weighted by
their respective probability to be observed.

We construct the likelihood function under the *working independence assumption*
[40, 41]. This means that for the time being we ignore the clustered structure of the
data caused by the correlation between granules within the same cell. Following the
work of Yu & Peng [41] on cure models, a particular type of discrete frailty model with
$\eta = 0$, we then integrate the frailty out to obtain a marginal likelihood function. The
resulting marginal *independence* log-likelihood, l_I, gives valid maximum likelihood

estimate (MLE) $\hat{\theta}_I$ of the parameter vector θ, but the inverse of the observed Hessian of   174
the independence log-likelihood, $\hat{H}^{-1}$, does not yield valid estimates of e.g. standard   175
errors [40, 42]. Thus, in order to construct confidence intervals or perform inferential   176
tests, corrections must be introduced.   177

Under the independence assumption, the log-likelihood, conditional on the frailty Z,   178
is given as the sum of the individual contributions,   179

$$\tilde{\ell}_I(\theta|t_{ij}, d_{ij}, X_{ij}, Z_{ij}) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \tilde{\ell}_{ij}(\theta|t_{ij}, d_{ij}, X_{ij}, Z_{ij}), \qquad (2)$$

with

$$\tilde{\ell}_{ij} = \log \left[\mu(t_{ij}|X_{ij}, Z_{ij})^{d_{ij}} e^{-M(t_{ij}|X_{ij}, Z_{ij})} \right],$$

where M is the cumulative hazard function,

$$M(t|X, Z) = \int_0^t \mu(s|X, Z)ds = Z \int_0^t \mu_0(s|X)ds = ZM_0(t|X),$$

and M_0 the cumulative baseline hazard. Since Z is unobservable, it must be integrated   180
out of the log-likelihood (2) to obtain the MLE. This procedure yields the marginal   181
(unconditional on Z) independence log-likelihood   182

$$\ell_I(\theta|t_{ij}, d_{ij}, X_{ij}) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \ell_{ij}(\theta|t_{ij}, d_{ij}, X_{ij}), \qquad (3)$$

where   183

$$\ell_{ij}(\theta|t_{ij}, d_{ij}, X_{ij}) =$$
$$\log \left[\mu_0(t_{ij}|X_{ij})^{d_{ij}} \left(\pi_{X_{ij}} \eta^{d_{ij}} e^{-\eta M_0(t_{ij}|X_{ij})} + (1 - \pi_{X_{ij}})e^{-M_0(t_{ij}|X_{ij})} \right) \right] \qquad (4)$$

is found by averaging the likelihood function with respect to Z [13]. Given the data, ℓ_I 184
can then be maximized to yield the MLE $\hat{\theta}_I$.   185

A commonly used approach to correct for clustering is to estimate the   186
variance-covariance matrix using the so-called *robust* or *sandwich* estimator [40]   187

$$\mathcal{R} = \hat{H}^{-1}\hat{V}\hat{H}^{-1}, \qquad \hat{V} = \sum_j U_j(\hat{\theta}_I)U_j(\hat{\theta}_I)', \qquad (5)$$

where

$$U_j(\theta) = \sum_i U_{ij}(\theta) = \sum_i \frac{\partial}{\partial \theta} \ell_{ij}(\theta|t_{ij}, d_{ij}, X_{ij})$$

is the score contribution from cluster j. From $\mathcal{R}$, robust standard errors for $\hat{\theta}_I$, $SE_{\hat{\theta}_I}$,   188
and correct Wald-type 95% confidence intervals, $\hat{\theta}_I \pm 1.96\, SE_{\hat{\theta}_I}$, can be obtained.   189
However, Wald tests are not reliable for testing null hypothesis with parameters on the   190
boundary of the parameter space (e.g., $\pi_X = 0$). Further, Wald-type inference can be   191
difficult to interpret when covariates are highly correlated, and inference based on the   192
likelihood ratio is preferable in finite samples [40].   193

In order to calculate valid likelihood-based confidence intervals, and perform   194
likelihood ratio tests, Chandler & Bate [40] proposed to adjust the independence   195
likelihood in order to obtain an *adjusted* log-likelihood function ℓ_A that has the same   196
MLE as ℓ_I, but has the 'correct' observed Hessian $\hat{H}_A$, i.e., the sandwich estimator in   197

(5) is obtained as the inverse of the observed Hessian, $\hat{H}_A = -\mathcal{R}^{-1}$. This can be obtained by defining [40]

$$\ell_A(\theta) = \ell_I\big(C(\theta - \hat{\theta}_I) + \hat{\theta}_I\big), \tag{6}$$

where $\hat{\theta}_I$ maximizes ℓ_I, and $C = N^{-1}N_A$ with $N'N = \hat{H}$ and $N'_A N_A = \hat{H}_A = -\mathcal{R}^{-1}$. The matrix square-roots N and N_A are conveniently constructed from the spectral decompositions of $\hat{H}$ and $-\mathcal{R}^{-1}$ [40]. Likelihood ratio tests, e.g. of the null hypothesis that a given parameter is equal to zero, say $\theta_k = 0$, can then be performed by comparing quantiles of the χ_1^2 distribution to the log likelihood ratio statistics $\Lambda_A = 2(\ell_A(\hat{\theta}_I) - \ell_A(\tilde{\theta}))$, where $\tilde{\theta}$ maximizes ℓ_A under the constraint $\theta_k = 0$. When performing tests against a null hypothesis with parameters on the boundary of the parameter space, e.g., testing $\pi_0 = 0$ or $\pi_1 = 0$, Λ_A is compared to the quantiles of the mixture distribution $(\chi_0^2 + \chi_1^2)/2$ [43]. This amounts to performing a one-sided test.

As recommended by Chandler & Bate [40], we chose a reparameterization that led to symmetric log-likelihoods, as verified by the symmetry of the estimated confidence intervals. In particular, we estimated the parameter vector $\theta = (\alpha_1, \alpha_2, \beta_2, \tilde{\eta}, \tilde{\pi}_0, \tilde{\pi}_1) = (\log \rho_1, \log \rho_2, \beta_2, \log \eta, \sqrt{\pi_0}, \sqrt{\pi_1})$, and calculated confidence intervals and performed log likelihood ratio tests for θ based on ℓ_A. For ease of interpretations, inferential results are presented for the original parameters using the inverse transformations of point and interval estimates. Calculations were performed in R [37]. Optimization of ℓ_I was done using the `nlminb` function. Numerical approximations to $\hat{H}$ and $\hat{U}_j$ were found using the `hessian` and `grad` functions from the numDeriv R package [44]. Spectral decompositions were obtained using the `eigen` R function.

Results

Time-varying, piecewise-constant hazard Poisson analysis

As explained in the Methods, we assume a piecewise constant hazard

$$\mu(t|X) = \begin{cases} \rho_0 e^{\beta_0 X} = e^{\alpha_0 + \beta_0 X}, & 0 \le t < 1 \text{ seconds}, \\ \rho_1 e^{\beta_1 X} = e^{\alpha_1 + \beta_1 X}, & s \le t < s+1 \text{ for some } s = 10, 20, 30, \dots \text{ seconds}, \\ \rho_2 e^{\beta_2 X} = e^{\alpha_2 + \beta_2 X}, & s+1 \le t < s+10 \text{ for some } s = 0, 10, 20, \dots \text{ seconds}, \end{cases}$$

where ρ_k and β_k model baseline hazards and effects of diabetes, respectively, and the subscripts $k = 0, 1, 2$ refer to, respectively, the first pulse, the following pulses and the intervals between pulses. In the Poisson formulation, the parameters $\alpha_k = \log(\rho_k)$ and β_k, $k = 0, 1, 2$ are estimated. However, to facilitate the interpretation of the baseline rates, inferential results are reported for ρ_k and β_k, $k = 0, 1, 2$.

The estimated rate related to the first pulse ($\hat{\rho}_0 = 0.176 \text{ s}^{-1}$) was found to be significantly greater (about 14-fold) than the estimate related to the other pulses ($\hat{\rho}_1 = 0.0013 \text{ s}^{-1}$; $p < 10^{-11}$). As expected, the estimated rate between pulses ($\hat{\rho}_2 = 0.0003 \text{ s}^{-1}$) was significantly lower (about 6-fold) than the rate during stimuli, reflecting that exocytosis mainly occur when Ca^{2+} channels open in response to the depolarizing K^+ pulses. Interestingly, diabetes had no statistically significant effect on the hazard during pulses, though there was a tendency towards a reduced rate ($\sim 75\%$ reduction) of exocytosis during the first pulse in diabetic cells ($p = 0.157$, $\hat{\beta}_0 = -1.48$, $\exp(\hat{\beta}_0) = 0.23$). This reduced rate was however poorly estimated as reflected by the large confidence interval. On the contrary, between pulses the rate of exocytosis was 2-3 fold *higher* in diabetic cells than in healthy cells ($p = 0.060$, $\hat{\beta}_2 = 0.98$, $\exp(\hat{\beta}_2) = 2.66$). We note that all the tests, except for β_1, (erroneously) show significance if clustering is

Table 1. **Estimated parameters using Poisson modeling.**

Parameter	estimate	95% CI	p-value	p-value (no clustering)
ρ_0	0.0176	(0.0040, 0.0781)	$< 10^{-11}$	$< 10^{-11}$
ρ_1	0.0013	(0.0004, 0.0040)		
ρ_2	0.0003	(0.0001, 0.0005)	0.006	$< 10^{-5}$
β_0	-1.48	(-3.54, 0.57)	0.157	0.046
β_1	-0.09	(-1.81 1.62)	0.914	0.858
β_2	0.98	(-0.04, 2.01)	0.060	0.0006

Wald-type 95% confidence intervals (CI) and p-values are based on the sandwich estimator $\mathcal{R}$ and t-tests. For β_k, the p-values refer to the null-hypotheses $\beta_k = 0$. For ρ_0 and ρ_2, the p-values refer to the null-hypotheses $\rho_0 = \rho_1$ and $\rho_2 = \rho_1$, respectively. The last column shows naive p-values ignoring clustering.

ignored. In summary, whereas healthy cells showed a prominent peak of exocytosis in response to the first pulse followed by bursts of release synchronized with the stimulating pulses, exocytosis occurred less well controlled by the stimuli in diabetic cells, as seen from nonparametric estimates of the cumulative incidence functions (Fig 2, black curves). These results correspond well to clinical characteristics of diabetes, where biphasic insulin secretion is disturbed [18].

## Frailty modeling of granule heterogeneity

The model presented in the previous subsection assume that all granules follow the same hazard function describing the rate of exocytosis. The obtained results suggest that this hazard declines from the first to subsequent stimulus pulses, and thus, that the peak of secretion is caused by a decrease in the rate of exocytosis. Alternatively, the peak of secretion is often attributed to a distinct *immediately releasable pool* (IRP) of granules that undergo exocytosis more rapidly that the non-IRP granules.

To account for a heterogeneous granule population, we introduced a discrete frailty variable Z that modeled the probability (π_X) of a granule to belong to the IRP and the fold-increase in exocytosis rate in the IRP via the parameter η. The size of the IRP is hence described by π_X, which was allowed to vary between healthy and diabetic cells. As described in the Methods, Z is a non-observable variable since we cannot *a priori* identify the granules that belong to the IRP.

In order to estimate the parameters in the model, we integrate Z out to obtain the marginal independence likelihood function ℓ_I, which is then maximized. The resulting maximum likelihood estimates of the parameters are reported in Table 2. The Hessian of the independence likelihood function ℓ_I does not provide valid estimates for standard errors since it neglects the clustered structure of the data caused by the correlation between granules within the same cell. In order to perform valid inference, we adjust ℓ_I to obtain the adjusted likelihood function ℓ_A, which permits us to construct confidence intervals from the log-likelihood statistics, and to perform ordinary log-likelihood ratio tests taking clustering into consideration [40] (Table 2).

In contrast to the Poisson model, we assumed that the rate of exocytosis, for a given granule (conditional on the frailty Z), was different during, compared to between, K^+ pulses, but that the hazard was independent of the pulse number, i.e., $\rho_0 = \rho_1$ (Fig 1). Based on the Poisson model, and after performing model selection in the frailty formulation, we assumed that diabetes did not influence the rate of exocytosis during pulses, i.e., $\beta_1 = 0$. However, diabetes was allowed to have an effect on the rate of exocytosis between pulses. As shown in Fig 2, the model provides a good overall fit to
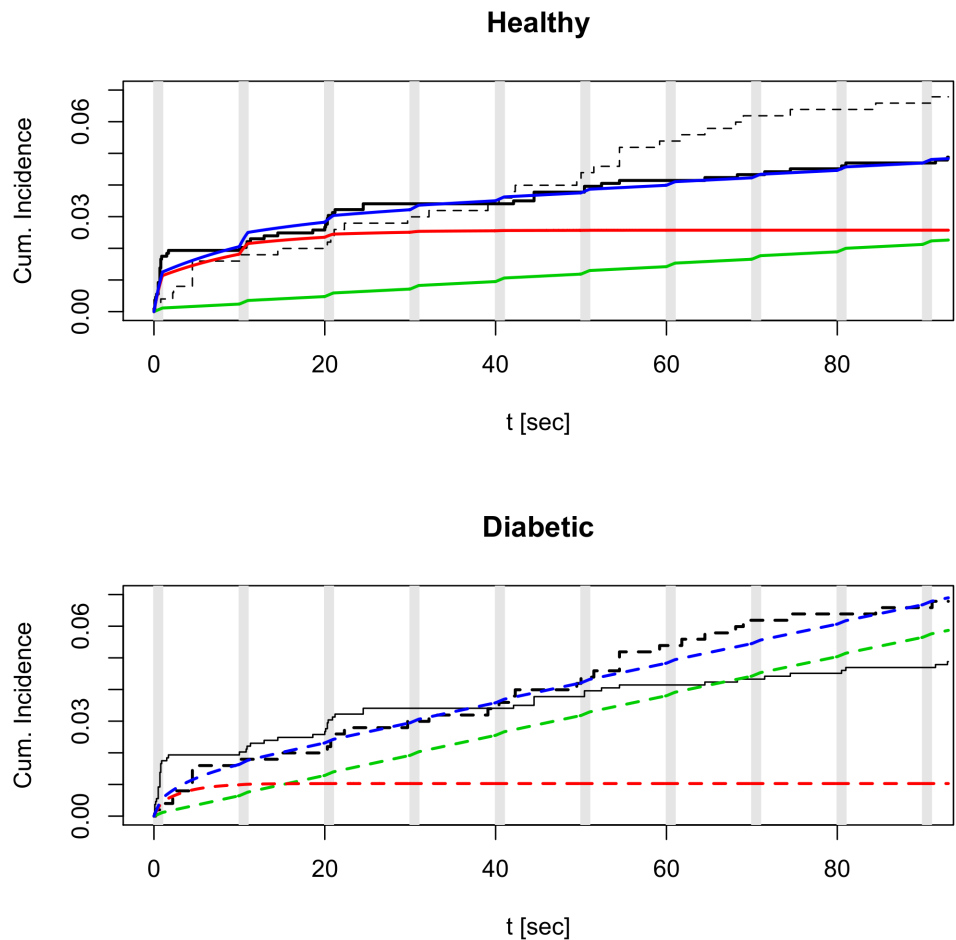
Fig 2. Estimated cumulative incidence probabilities. The curves represent the estimated probabilities of an exocytotic event before time t (the cumulative incidence) for a given granule in healthy (upper panel, full curves) or diabetic (lower panel, dashed curves) β-cells. The black curves are obtained from model-free, non-parametric Kaplan-Meier estimates, which, for comparison, are shown in both panels. Steps in these curves correspond to exocytotic events. For the frailty model we show the marginal estimate (blue), and the estimates conditional on the frailty, $Z = \eta$ (IRP granules; red; scaled by π_X) or $Z = 1$ (non-IRP granules; green; scaled by $1 - \pi_X$). The gray vertical lines indicate the K^+ pulses.

the data.   275

Our results (Table 2) concerning the estimated frailty parameter $\hat{\pi}_0$ suggest that the   276
size of the IRP in healthy β-cells is significantly greater that zero ($p = 0.022$) and   277
amounts to $\sim 2.6\%$ of the docked granules. Note that if clustering is (erroneously)   278
neglected, the significance of the test becomes extremely high ($p < 10^{-7}$). In contrast,   279
in diabetic β-cells, the estimated IRP size, as measured by $\hat{\pi}_1$, is only $\sim 1.0\%$ of the 280
docked granules, almost significantly different from zero ($p = 0.052$, 95% C.I.   281
(0.00003,0.035)). Again, if clustering is neglected the difference becomes (erroneously)   282
highly significant ($p = 0.008$). Thus, we reiterate that in order to perform correct   283
inference, clustering must be taken into account.   284

IRP granules have a $\sim$500-fold higher rate of exoytosis ($\hat{\eta} = 499.5$, $p < 10^{-4}$) compared to non-IRP granules, i.e., the rate-of-exocytosis during pulses was estimated to $\hat{\rho}_1 = 0.0012$ s^{-1} in non-IRP granules and to $\hat{\eta}\hat{\rho}_1 = 0.5837$ s^{-1} in IRP granules. The estimated between-pulse rate in healthy cells $\hat{\rho}_2 = 0.00014$ s^{-1} was $\sim$8 times lower than $\hat{\rho}_1$ ($p = 0.0005$). Note that these estimates compare well with the Poisson model results reported in Table 1. Interestingly, and in line with the Poisson modeling, exocytosis between pulses was estimated to occur at a higher rate in diabetic cells ($\hat{\rho}_2 e^{\hat{\beta}_2} = 6.0 \cdot 10^{-4}$ s^{-1} in diabetic cells vs. $\hat{\rho}_2 = 1.4 \cdot 10^{-4}$ s^{-1} in healthy cells, $p = 0.038$). As typically seen for covariate effects [13], the effect of diabetes on the between-pulse hazard was estimated to be greater in the frailty formulation compared to the Poisson model without frailty.

Table 2. Estimated parameters using frailty modeling.

Parameter	MLE	95% CI	p-value	p-value (no clustering)
ρ_1	0.00117	(0.00079, 0.00174)		
ρ_2	0.00014	(0.00011, 0.00019)	0.0005	$4 \cdot 10^{-7}$
β_2	1.43	(0.98, 1.84)	0.038	0.0002
η	499.5	(321.7, 773.4)	< 0.0001	$< 10^{-7}$
π_0	0.026	(0.013, 0.044)	0.022	$< 10^{-7}$
π_1	0.010	(0.00003, 0.035)	0.052	0.008

Maximum likelihood estimates (MLE) are based on the independence likelihood function ℓ_I. The 95% confidence intervals (CI) and tests of hypotheses are based on the log likelihood ratio statistic Λ_A obtained from the adjusted log-likelihood function ℓ_A. For ρ_2, the p-value refers to the null hypothesis $\rho_1 = \rho_2$. The last column shows p-values based on log likelihood ratio test using ℓ_I ignoring clustering.

Discussion

The aim of this paper was to present a novel application of a well-established statistical methodology to modern cell biological data obtained with live cell imaging. To the best of our knowledge, a rigorous and statistically sound method for the analysis of exocytosis data obtained by TIRF microscopy has been lacking.

The presented method can take into account unobserved heterogeneity by the inclusion of frailties, here exemplified by a discrete frailty representing the IRP. In addition, observed covariates, here whether a cell came from a healthy or diabetic donor, can be included for example in a proportional hazards formulation. We envisage that our approach to the study of exocytosis with the use of flexible survival modeling [12] can be extended to include more complicated, time-dependent covariates [45], such as for example Ca^{2+} concentrations [46, 47] or protein levels [5, 7, 48] at the granules. Further extensions could take into consideration spatial information in addition to the temporal data [49]. The current formulation can also readily handle more complex censoring patterns than the pattern considered here, such as for example experiments interrupted at different times. Further, the method is not limited to the study of exocytosis or to endocrine cells; TIRF imaging of exocytotic events of e.g. synaptic vesicles [8] or GLUT4 vesicles in fat or muscle cells [50, 51], or of individual endocytotic events [50, 52], produce data similar to the dataset analyzed here. Moreover, the statistical methodology was here applied to data from TIRF imaging, but it is suitable for analyzing well-defined temporal cellular events recorded with any other imaging technique.

It is a well-known fact in the statistical literature [40, 42], but often not considered in biology, that ignoring clustering typically leads to underestimation of standard errors, and thus to small 'naive' p-values. Our example shows clearly this effect, which is due to the correlation between granules in the same cell: some cells are inherently 'highly responding', meaning that the granules in such a cell readily undergo exocytosis, while other cells are not. Ignoring this fact, would lead, for example, to rejecting the null hypotheses $\pi_1 = 0$ (Table 2) or, in the Poisson formulation, $\beta_0 = 0$ (Table 1). Taking into account the clustered structure of the data yields more cautious conclusions.

Our study also highlights how different statistical models can explain the data, but with different biological interpretations. The Poisson formulation assumed that all granules in a cell behave similarly, but that the rate of exocytosis is higher during the first pulse compared to the subsequent pulses. In contrast, the frailty model assumes that the rate of exocytosis for a given granule is the same in all pulses, but that the granule population is heterogeneous, since some granules belong to the IRP and have higher exocytosis rate. The latter model respects better various biological results regarding exocytosis in β-cells. Thus, for this kind of studies of complex cell biological questions, a close interaction between biologists and statisticians is needed in order to formulate a biologically correct model, which then serves as the basis for performing statistical inference with results that are both biologically reasonable and statistically sound.

Our application of the frailty model to human β-cells estimated that the IRP constitutes 2-3% of the docked granules in healthy cells, but only approximately half as many in diabetic cells. In diabetic cells, we were unable to conclude whether an IRP is present; the estimate of π_1 was borderline significant ($p = 0.052$). Further studies should investigate this aspect further. Based on a cell capacitance of ~10 pF [53], an absolute membrane capacitance of 10 fF/μm^2, and assuming a density of ~0.8 docked granule per μm^2 membrane [54], the number of docked granules can be estimated to be ~800/cell. Hence, we estimate that the IRP contains ~20 granules in healthy cells, and around 10 granules in diabetic cells. The estimate in healthy cells corresponds well to the estimate of the IRP in unstimulated mouse β-cells [55].

The formulation of the model assumed piecewise constant baseline hazard. This formulation allowed us to perform explicit maximum-likelihood estimation, and to quantify the rate of exocytosis during and between pulses. As expected, we found that the rate of exocytosis was higher during pulses, compared to during the interval between pulses where Ca^{2+} channels are closed. This suggests that in healthy β-cells a close coupling between Ca^{2+} channels and insulin granules guarantees tight control of synchronized secretion. Interestingly, between pulses the rate of exocytosis was significantly higher in diabetic cells, as was the total amount of exocytosis during the experiments (Fig 2). This asynchronous release may correspond to basal insulin secretion, which is increased in diabetic mouse models [22], and even in early phases of human diabetes [56, 57], in agreement with our findings. The higher between-pulse rate might be explained by a looser coupling between Ca^{2+} channels and insulin granules in diabetic cells [22], so that residual Ca^{2+} remaining after the end of the K$^+$ pulse and closure of Ca^{2+} channels triggers unsynchronized exocytosis. Such a scenario would require that the Ca^{2+} affinity for exocytosis is higher in the granules located away from the Ca^{2+} channels [58].

In summary, we have shown how to adapt time-to-event analysis to the study of TIRF imaging data of exocytosis in human β-cells. This powerful statistical methodology allows quantifying several biologically interesting parameters, such as rates of exocytosis, probabilities of an event in a certain time interval, and the size of the IRP, in healthy and diabetic β-cells. In this context, rigorous statistical tests taking into consideration the clustered structure of the data are needed to reflect the correlation

between granules within the same cell. This makes it possible to correctly investigate [370] hypotheses of disturbances in diseased cells. We believe the presented approach, which [371] should be seen as a starting point for future extensions, could be generally applicable to [372] analysis of a range of cell biological data with well-defined temporal events, also in the [373] presence of more complicated covariates and censoring patterns. [374]

References

1. Lang T, Wacker I, Steyer J, Kaether C, Wunderlich I, Soldati T, et al. Ca2+-triggered peptide secretion in single cells imaged with green fluorescent protein and evanescent-wave microscopy. Neuron. 1997;18(6):857–63.

2. Steyer JA, Horstmann H, Almers W. Transport, docking and exocytosis of single secretory granules in live chromaffin cells. Nature. 1997;388(6641):474–8. doi:10.1038/41329.

3. Ohara-Imaizumi M, Nakamichi Y, Tanaka T, Ishida H, Nagamatsu S. Imaging exocytosis of single insulin secretory granules with evanescent wave microscopy: distinct behavior of granule motion in biphasic insulin release. J Biol Chem. 2002;277(6):3805–8. doi:10.1074/jbc.C100712200.

4. Michael DJ, Xiong W, Geng X, Drain P, Chow RH. Human insulin vesicle dynamics during pulsatile secretion. Diabetes. 2007;56(5):1277–88. doi:10.2337/db06-0367.

5. Barg S, Knowles MK, Chen X, Midorikawa M, Almers W. Syntaxin clusters assemble reversibly at sites of secretory granules in live cells. Proc Natl Acad Sci U S A. 2010;107(48):20804–9. doi:10.1073/pnas.1014823107.

6. Toomre D. Generating live cell data using total internal reflection fluorescence microscopy. Cold Spring Harb Protoc. 2012;2012(4):439–46. doi:10.1101/pdb.ip068676.

7. Gandasi NR, Barg S. Contact-induced clustering of syntaxin and munc18 docks secretory granules at the exocytosis site. Nat Commun. 2014;5:3914. doi:10.1038/ncomms4914.

8. Midorikawa M, Sakaba T. Imaging Exocytosis of Single Synaptic Vesicles at a Fast CNS Presynaptic Terminal. Neuron. 2015;88(3):492–8. doi:10.1016/j.neuron.2015.09.047.

9. Burgoyne RD, Morgan A. Secretory granule exocytosis. Physiol Rev. 2003;83(2):581–632. doi:10.1152/physrev.00031.2002.

10. Karatekin E, Tran VS, Huet S, Fanget I, Cribier S, Henry JP. A 20-nm step toward the cell membrane preceding exocytosis may correspond to docking of tethered granules. Biophys J. 2008;94(7):2891–905. doi:10.1529/biophysj.107.116756.

11. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media; 2005.

12. Cortese G, Scheike TH, Martinussen T. Flexible survival regression modelling. Stat Methods Med Res. 2010;19(1):5–28. doi:10.1177/0962280209105022.

13. Wienke A. Frailty models in survival analysis. CRC Press; 2010.

14. Kahn SE. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes. Diabetologia. 2003;46(1):3–19. doi:10.1007/s00125-002-1009-0.

15. Ashcroft FM, Rorsman P. Diabetes mellitus and the β cell: the last ten years. Cell. 2012;148(6):1160–71. doi:10.1016/j.cell.2012.02.010.

16. Rorsman P, Braun M. Regulation of insulin secretion in human pancreatic islets. Annu Rev Physiol. 2013;75:155–79. doi:10.1146/annurev-physiol-030212-183754.

17. Curry DL, Bennett LL, Grodsky GM. Dynamics of insulin secretion by the perfused rat pancreas. Endocrinology. 1968;83(3):572–584.

18. Del Prato S, Marchetti P, Bonadonna RC. Phasic insulin release and metabolic regulation in type 2 diabetes. Diabetes. 2002;51 Suppl 1:S109–S116.

19. Del Guerra S, Lupi R, Marselli L, Masini M, Bugliani M, Sbrana S, et al. Functional and molecular defects of pancreatic islets in human type 2 diabetes. Diabetes. 2005;54(3):727–735.

20. Daniel S, Noda M, Straub SG, Sharp GW. Identification of the docked granule pool responsible for the first phase of glucose-stimulated insulin secretion. Diabetes. 1999;48(9):1686–1690.

21. Rorsman P, Renström E. Insulin granule dynamics in pancreatic beta cells. Diabetologia. 2003;46(8):1029–1045. doi:10.1007/s00125-003-1153-1.

22. Collins SC, Hoppa MB, Walker JN, Amisten S, Abdulkader F, Bengtsson M, et al. Progression of diet-induced diabetes in C57BL6J mice involves functional dissociation of Ca2(+) channels from secretory vesicles. Diabetes. 2010;59(5):1192–201. doi:10.2337/db09-0791.

23. Grodsky GM. A threshold distribution hypothesis for packet storage of insulin and its mathematical modeling. J Clin Invest. 1972;51:2047–2059. doi:10.1172/JCI107011.

24. Thomas P, Wong JG, Lee AK, Almers W. A low affinity Ca2+ receptor controls the final steps in peptide secretion from pituitary melanotrophs. Neuron. 1993;11(1):93–104.

25. Heinemann C, Chow RH, Neher E, Zucker RS. Kinetics of the secretory response in bovine chromaffin cells following flash photolysis of caged Ca2+. Biophys J. 1994;67(6):2546–2557. doi:10.1016/S0006-3495(94)80744-1.

26. Barg S, Ma X, Eliasson L, Galvanovskis J, Göpel SO, Obermüller S, et al. Fast exocytosis with few Ca(2+) channels in insulin-secreting mouse pancreatic B cells. Biophys J. 2001;81(6):3308–3323.

27. Moser T, Neher E. Rapid exocytosis in single chromaffin cells recorded from mouse adrenal slices. J Neurosci. 1997;17(7):2314–2323.

28. Pedersen MG. On depolarization-evoked exocytosis as a function of calcium entry: possibilities and pitfalls. Biophys J. 2011;101(4):793–802. doi:10.1016/j.bpj.2011.06.061.

29. Voets T. Dissection of three Ca2+-dependent steps leading to secretion in chromaffin cells from mouse adrenal slices. Neuron. 2000;28(2):537–545.

30. Bertuzzi A, Salinari S, Mingrone G. Insulin granule trafficking in beta-cells: mathematical model of glucose-induced insulin secretion. Am J Physiol Endocrinol Metab. 2007;293(1):E396–E409. doi:10.1152/ajpendo.00647.2006.

31. Pedersen MG, Corradin A, Toffolo GM, Cobelli C. A subcellular model of glucose-stimulated pancreatic insulin secretion. Philos Transact A Math Phys Eng Sci. 2008;366(1880):3525–3543. doi:10.1098/rsta.2008.0120.

32. Pedersen MG, Sherman A. Newcomer insulin secretory granules as a highly calcium-sensitive pool. Proc Natl Acad Sci U S A. 2009;106(18):7432–7436. doi:10.1073/pnas.0901202106.

33. Montefusco F, Pedersen MG. Mathematical modelling of local calcium and regulated exocytosis during inhibition and stimulation of glucagon secretion from pancreatic alpha-cells. J Physiol. 2015;593(20):4519–30. doi:10.1113/JP270777.

34. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 1979;16(3):439–54.

35. Gandasi NR, Vestö K, Helou M, Yin P, Saras J, Barg S. Survey of Red Fluorescence Proteins as Markers for Secretory Granule Exocytosis. PLoS One. 2015;10(6):e0127801. doi:10.1371/journal.pone.0127801.

36. Dalgaard P. Introductory statistics with R. 2nd ed. New York: Springer; 2008.

37. R Core Team. R: A Language and Environment for Statistical Computing; 2016. Available from: `https://www.R-project.org/`.

38. Arai M. Cluster-robust standard errors using R; 2015. Available from: `http://www.ne.su.se/polopoly_fs/1.216115.1426234213!/menu/standard/file/clustering1.pdf`.

39. Rorsman P, Eliasson L, Kanno T, Zhang Q, Gopel S. Electrophysiology of pancreatic β-cells in intact mouse islets of Langerhans. Prog Biophys Mol Biol. 2011;107(2):224–35. doi:10.1016/j.pbiomolbio.2011.06.009.

40. Chandler RE, Bate S. Inference for clustered data using the independence loglikelihood. Biometrika. 2007;94(1):167–183.

41. Yu B, Peng Y. Mixture cure models for multivariate survival data. Computational Statistics & Data Analysis. 2008;52(3):1524–1532.

42. Davison AC. Statistical models. Cambridge series in statistical and probabilistic mathematics. Cambridge, U.K.: Cambridge University Press; 2003. Available from: `http://www.loc.gov/catdir/description/cam041/2003279768.html`.

43. Claeskens G, Nguti R, Janssen P. One-sided tests in shared frailty models. Test. 2008;17(1):69–82.

44. Gilbert P, Varadhan R. numDeriv: Accurate Numerical Derivatives; 2016. Available from: `https://CRAN.R-project.org/package=numDeriv`.

45. Cortese G, Andersen PK. Competing risks and time-dependent covariates. Biom J. 2010;52(1):138–58. doi:10.1002/bimj.200900076.

46. Becherer U, Moser T, Stühmer W, Oheim M. Calcium regulates exocytosis at the level of single vesicles. Nat Neurosci. 2003;6(8):846–53. doi:10.1038/nn1087.

47. Hoppa MB, Collins S, Ramracheya R, Hodson L, Amisten S, Zhang Q, et al. Chronic palmitate exposure inhibits insulin secretion by dissociation of Ca(2+) channels from secretory granules. Cell Metab. 2009;10(6):455–65. doi:10.1016/j.cmet.2009.09.011.

48. Trexler AJ, Sochacki KA, Taraska JW. Imaging the recruitment and loss of proteins and lipids at single sites of calcium-triggered exocytosis. Mol Biol Cell. 2016;27(15):2423–34. doi:10.1091/mbc.E16-01-0057.

49. Rúa SMH, Dey DK. A transformation class for spatio-temporal survival data with a cure fraction. Statistical methods in medical research. 2016;25(1):167–187.

50. Stenkula KG, Lizunov VA, Cushman SW, Zimmerberg J. Insulin controls the spatial distribution of GLUT4 on the cell surface through regulation of its postfusion dispersal. Cell Metab. 2010;12(3):250–9. doi:10.1016/j.cmet.2010.08.005.

51. Lizunov VA, Lee JP, Skarulis MC, Zimmerberg J, Cushman SW, Stenkula KG. Impaired tethering and fusion of GLUT4 vesicles in insulin-resistant human adipose cells. Diabetes. 2013;62(9):3114–9. doi:10.2337/db12-1741.

52. Stabley DR, Oh T, Simon SM, Mattheyses AL, Salaita K. Real-time fluorescence imaging with 20?nm axial resolution. Nat Commun. 2015;6:8307. doi:10.1038/ncomms9307.

53. Braun M, Ramracheya R, Bengtsson M, Zhang Q, Karanauskaite J, Partridge C, et al. Voltage-gated ion channels in human pancreatic beta-cells: electrophysiological characterization and role in insulin secretion. Diabetes. 2008;57(6):1618–1628. doi:10.2337/db07-0991.

54. Edlund A, Esguerra JLS, Wendt A, Flodström-Tullberg M, Eliasson L. CFTR and Anoctamin 1 (ANO1) contribute to cAMP amplified exocytosis and insulin secretion in human and murine pancreatic beta-cells. BMC Med. 2014;12:87. doi:10.1186/1741-7015-12-87.

55. Eliasson L, Abdulkader F, Braun M, Galvanovskis J, Hoppa MB, Rorsman P. Novel aspects of the molecular mechanisms controlling insulin secretion. J Physiol. 2008;586(14):3313–3324. doi:10.1113/jphysiol.2008.155317.

56. Corkey BE. Banting lecture 2011: hyperinsulinemia: cause or consequence? Diabetes. 2012;61(1):4–13. doi:10.2337/db11-1483.

57. Pories WJ, Dohm GL. Diabetes: have we got it all wrong? Hyperinsulinism as the culprit: surgery provides the evidence. Diabetes Care. 2012;35(12):2438–42. doi:10.2337/dc12-0684.

58. Pedersen MG. Insulin secretory granules enter a highly calcium-sensitive state following palmitate-induced dissociation from calcium channels: a theoretical study. J Neuroendocrinol. 2010;22(12):1315–1324. doi:10.1111/j.1365-2826.2010.02056.x.