Assessing Cross-Contamination in Spike-Sorted Electrophysiology Data

Jack P. Vincent[1,2] and Michael N. Economo[1,2,3]

[1]Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, [2]Center for Neurophotonics, Boston University, Boston, Massachusetts 02215, and [3]Center for Systems Neuroscience, Boston University, Boston, Massachusetts 02215

Abstract

Recent advances in extracellular electrophysiology now facilitate the recording of spikes from hundreds or thousands of neurons simultaneously. This has necessitated both the development of new computational methods for spike sorting and better methods to determine spike-sorting accuracy. One long-standing method of assessing the false discovery rate (FDR) of spike sorting—the rate at which spikes are assigned to the wrong cluster—has been the rate of interspike interval (ISI) violations. Despite their near ubiquitous usage in spike sorting, our understanding of how exactly ISI violations relate to FDR, as well as best practices for using ISI violations as a quality metric, remains limited. Here, we describe an analytical solution that can be used to predict FDR from the ISI violation rate (ISI_v). We test this model in silico through Monte Carlo simulation and apply it to publicly available spike-sorted electrophysiology datasets. We find that the relationship between ISI_v and FDR is highly nonlinear, with additional dependencies on firing frequency, the correlation in activity between neurons, and contaminant neuron count. Predicted median FDRs in public datasets recorded in mice were found to range from 3.1 to 50.0%. We found that stochasticity in the occurrence of ISI violations as well as uncertainty in cluster-specific parameters make it difficult to predict FDR for single clusters with high confidence but that FDR can be estimated accurately across a population of clusters. Our findings will help the growing community of researchers using extracellular electrophysiology assess spike-sorting accuracy in a principled manner.

Key words: electrophysiology; spike sorting

Significance Statement

High-density silicon probes are widely used to record the activity of large populations of neurons while animals are engaged in complex behavior. In this approach, each electrode records spikes from many neurons, and "spike-sorting" algorithms are used to group the spikes originating from each neuron together. This process is error-prone, however, and so the ability to assess spike-sorting accuracy is essential for properly interpreting neural activity. The rate of interspike interval (ISI) violations is commonly used to assess spike-sorting accuracy, but the relationship between the ISI violation rate (ISI_v) and sorting accuracy is complex and poorly understood. Here, we describe this relationship in detail and provide guidelines for how to properly use ISI_v to assess spike-sorting accuracy.

Introduction

Extracellular electrophysiology has become an increasingly popular method for studying neuronal activity at the population level. Silicon probes containing dozens or hundreds of densely packed electrode sites can be used to observe neuronal action potentials in many neurons simultaneously. This multiplexed signal acquisition, however, often necessitates the assignment of observed action potentials to individual neurons—"spike sorting"—as a critical first step prior to many analyses (Todorova et al., 2014; Rey et al., 2015). Approaches for spike sorting vary, though all generally involve comparisons of

observed action potential waveforms within and across electrodes and then grouping similar spikes—thought to be produced by the same neuron—together, forming clusters (Gibson et al., 2012; Quiroga and Panzeri, 2013). Ideally, each cluster is composed principally of true-positive (TP) spikes from a single neuron making it a "well-isolated" cluster. Spikes from this single neuron can be erroneously excluded from the cluster, resulting in false negatives (FNs). The primary focus of this work, however, is false positives (FPs), wherein spikes are misassigned to a cluster whose activity is meant to correspond to a different neuron. Clusters with substantial "contamination" by FPs thus represent the combined activity of multiple neurons.

FPs are a persistent problem in spike sorting that results from frequently unavoidable similarities in action potential waveforms, occurrence of spikes at overlapping times, nonstationarity in waveform shape, and recording noise. Contamination can distort the activity of a cluster, potentially leading to incorrect conclusions about how single neurons encode information. For instance, a population of neurons recorded during a task with two cues may contain neurons responsive to just one cue or the other. However, poor sorting may lead to cluster cross-contamination between these two phenotypes, giving the impression that neurons in this region respond to both cues. The prevalence of FPs in a single cluster, recording session, or dataset can be described by the false discovery rate (FDR), a value that ranges between 0 and 1 and reports the proportion of sorted spikes that have been misassigned. While FNs can also be a concern, as they can decrease recorded spike frequencies and thus reduce statistical power in subsequent analyses (Hill et al., 2011), FNs should not generally alter the overall patterns of activity associated with individual clusters and are thus of less concern than FPs.

Algorithmic approaches to spike sorting for high-density silicon probe electrophysiology have seen concentrated development over the last decade, with researchers now selecting from a number of competing options for high-throughput automated spike sorting (Bestel et al., 2012; Pachitariu et al., 2016; Chung et al., 2017; J. H. Lee et al., 2017; Jun et al., 2017; Chaure et al., 2018; Buccino et al., 2020; Saif-ur-Rehman et al., 2021; Toosi et al., 2021), yet concomitant techniques for post hoc analysis of spike-sorted data (Pouzat et al., 2002; Hill et al., 2011; Neymotin et al., 2011; Barnett et al., 2016; Magland et al., 2020) have been given comparatively less attention. Sorting quality metrics generally fall into two categories: assessment of cluster overlap using dimensionally reduced representations of sorted spike waveforms, e.g., L-ratio (Schmitzer-Torbert et al., 2005), isolation distance (Harris et al., 2001), D-prime (Hill et al., 2011), and silhouette score, or empirical measures known to be related to cluster isolation and sorting difficulty, e.g., signal-to-noise ratio, presence ratio, firing range, and interspike interval (ISI) violations. Among all these metrics, ISI violations are unique in that they are tightly linked to the occurrence of FPs. Biophysical limitations prevent neurons from producing consecutive spikes within their absolute refractory period, meaning the presence of action potentials spaced by less than the absolute refractory period, an ISI violation, is always the result of at least one FP.

ISI violations are typically reported as a fraction of the total number of spikes assigned to a cluster. The ISI violation rate (ISI_v)—the number of ISI violations divided by the total number of spikes assigned to a cluster—is often interpreted subjectively with only a general understanding that a lower ISI_v is associated with a lower FDR. Often, although not always, it is appreciated that most FP spikes do not produce ISI violations and so $ISI_v \ll FDR$. Previous work has estimated the relationship between ISI_v and FDR (Hill et al., 2011; Llobet et al., 2022) using simplifying assumptions, but the accuracy and limitations of predicting cluster FDR on the basis of ISI violations under realistic experimental conditions have not been assessed.

Here, we develop a comprehensive model explaining the relationship between ISI_v and FDR following spike sorting with respect to cluster contamination, neuronal firing frequency, the temporal relationships between neurons, and the number of neurons contributing FPs. We benchmark the accuracy of this model in silico through Monte Carlo simulation and explore limitations in the accuracy of FDR estimation imposed by in vivo recording conditions. Finally, we apply this model to publicly available spike-sorted electrophysiology data to estimate FDRs from ISI violations in published studies and provide researchers with intuition about the range of FDRs expected in silicon probe electrophysiology.

Materials and Methods

Monte Carlo simulation of neural spike trains. Stochastic neural spike trains were simulated using Elephant (Electrophysiology Analysis Toolkit; Yegenoglu et al., 2018). Specifically, neurons were modeled as either homogeneous or inhomogeneous Poisson processes (van Vreeswijk, 2010) using either the StationaryPoissonProcess or the NonStationaryPoissonProcess functions of the spike_train_generation module. Custom Python scripts were used for subsequent simulation and analysis. While Poisson point processes were used for all simulations of neural spike trains, post hoc custom simulations of gamma, inverse Gaussian, and log-normal point processes indicated that observed results did not seem to depend on any particular point process or ISI coefficient of variation (CV; Extended Data Fig. 3-1). A refractory period of 2.5 ms was assumed for all simulations as well as all calculations in Table 2. All datasets examined were recorded in mice of both sexes, and most recordings targeted the neocortex. However, for data collected from different brain regions or species, this parameter would need to be adjusted accordingly. Simulating "infinite" contaminant neurons was accomplished by simulating a single contaminant neuron with no refractory period. Simulated recording durations varied depending on the desired level of certainty in ISI_v and need to emulate realistic recording conditions. These times included ∼28 h (Fig. 3*A,B*), ∼17 h (Fig. 3*C*), 12 h (Fig. 3*D*), 30 min (Fig. 2*B*), and 10 min (Figs. 2*A,C*, 4*B,C*).

Table 1. Symbols used

Symbol	Description	Units
FDR	False discovery rate	Unitless or %
ISI_v	Interspike interval violation rate	Unitless or %
f_t	Total observed firing frequency	Hz
f_{TP}	Frequency of TP spikes	Hz
f_{FP}	Frequency of FP spikes	Hz
f_v	Frequency of ISI violations	Hz
τ	Neuronal refractory period	s
τ_c	Spike-sorting censor period	s
τ_e	Effective refractory period	s
N	Number of contaminant neurons	-
n	Number of elements in a firing frequency vector	-

Monte Carlo simulation of cluster populations. In some simulations, all parameters were varied simultaneously to produce populations of clusters with a wide range of physiologically feasible parametric combinations (Figs. 3D, 4B,C). For each cluster, the following multistep process was used: (1) FDR, N, and f_t were randomly sampled from bounded probability distributions. Specifically, FDR was sampled from bounded Cauchy distributions as these were found to most accurately replicate ISI_v distributions found in electrophysiological data when simulated. Different population median and mean FDRs were obtained by adjusting the location and scale of the Cauchy distribution. f_t was randomly selected from a bounded uniform distribution: [4, 20] (Fig. 3D), [1, 10] (Fig. 4B), and [4, 16] (Fig. 4C). Possible values of N were uniformly chosen from among the following values: [1, 2, 5, ∞]. This was found to be the most balanced way to sample N, as the effect of contaminant neuron count on ISI violation occurrence increases logarithmically. (2) Unit vector peristimulus time histograms (PSTHs) were randomly selected from one of the datasets in Table 2 to serve as $\hat{f}_{TP}$ and $\hat{f}_{FP}$. These were selected uniformly from among all clusters present in the dataset. (3) These PSTHs were then scaled appropriately and combined to reach the desired f_t and FDR. The correct scaling was determined using a function optimization routine (minimize_scalar, SciPy) that sought to minimize the difference between each scaled PSTH's current and desired average firing frequency. TP–FP covariance obtained using this method varied from −49.2 to 113.7 Hz^2. Longer simulated recording times were attained through repeated simulation of these PSTHs.

Calculation of PSTHs. PSTHs of cluster responses for in vivo electrophysiology data were calculated using bin sizes of 50 ms. Predicted median and mean FDR were found to be largely unaffected by bin size selection in trial-averaged PSTHs. For continuously recorded data, the greatest length of time that could be extracted around each cue without overlapping between "trials" was used to generate trial-averaged PSTHs.

Estimating unknown cluster parameters. When predicting FDR using recorded data, only limited information is available about each sorted cluster. $\vec{f}_t$ and ISI_v can be computed directly. τ can be obtained through prior knowledge of electrophysiological properties of neurons in the organism and brain region being recorded from. $\hat{f}_{FP}$ can be estimated through examination of other sorted clusters present in the recording session. Estimating N is not straightforward.

To account for uncertainty in N, two approaches can be taken. A reasonable guess (e.g., 2, 3) can be chosen, and simply plugged directly into the equation, or multiple values representing extreme cases can be input and their results averaged. For example, FDR can be calculated by assuming N equals either 1 or ∞ and then taking the average of results from these two cases. Empirically, we found this equivalent to assuming a single N of ~2–3. Our results indicate that FDR estimates are not highly sensitive to the particular choice of N (Fig. 2C).

Depending on the assumed N, $\hat{f}_{FP}$ can be estimated from either other single clusters or averaged combinations of clusters. For example, if an N of 1 is assumed, $\hat{f}_{FP}$ can be directly obtained from other random clusters, while if an N of 2 is assumed, $\hat{f}_{FP}$ can be obtained by taking the average of 2 other random clusters. If an N of ∞ is assumed, $\hat{f}_{FP}$ can be obtained from a global average PSTH. Clusters for deriving $\hat{f}_{FP}$ can be restricted to those on the same or nearby electrode sites.

Estimation of FDR in simulated cluster populations and electrophysiology datasets. For FDR predictions in both simulated cluster populations (Figs. 3D, 4B,C) and electrophysiology datasets (Table 2), the following methodology was used. An N of both 1 and ∞ were assumed: in the former case, a given sorted cluster was compared with every other cluster in the simulation or recording session, while in the latter, it was compared with a single global $\hat{f}_{FP}$. The final FDR is the mean of the ($N = 1$) cases averaged with the ($N = \infty$) case.

For estimating FDR in electrophysiology datasets (Table 2), clusters for deriving $\hat{f}_{FP}$ were not restricted to those on the same or nearby electrode sites due to uncertainty about each dataset's probe geometry. Probe geometry was also not implemented in simulations of cluster populations (Figs. 3D, 4B,C). A single τ of 2.5 ms was used for every dataset, all of which were comprised of recordings from mouse brains. This refractory period was modified by the censor period,

when necessary. Censor periods were determined through visual inspection of aggregated per-cluster ISI histograms across the entire dataset. For selecting datasets to examine, only papers published in the last 10 years with publicly available spike-sorted electrophysiology data were considered. No limits or minimums were placed on cluster count, and no sorting methods were specifically included or excluded. In some datasets, spontaneous activity was recorded without a cue to calculate PSTHs around. In such cases, FDR was initially predicted using firing frequency vectors calculated across the entire session. However, these predictions were found to be sensitive to the bin size used; therefore the predicted FDR assuming homogeneous firing is given instead (Table 1; Eq. 9).

Theoretical limits of predicted FDR. Due to stochasticity associated with ISI_v estimates, observed ISI_v values sometimes exceed theoretical bounds given by the chosen parameters, producing imaginary predicted FDRs. In such cases, the predicted FDR is capped at its theoretical maximum for a given (assumed) number of contaminant neurons (Eq. 1). This maximum is derived from the fact that given a certain N, if predicted FDR exceeds FDR_{max}, a lower FDR could be attained by simply selecting a different neuron within the sorted cluster to be the TP neuron. When FDR is being calculated by averaging the ($N = 1$) and ($N = \infty$) cases, FDR_{max} is the average of these two cases' maximums, 0.75, as follows:

$$FDR_{max} = \frac{N}{N+1}.$$ (1)

Code availability. The code for calculating FDR with user-provided spike–sorted electrophysiology data, reproducing figures, and simulating neuronal firing and spike sorting is available via GitHub (https://github.com/economolab/DCISIv).

Results

The relationship between ISI_v and FDR is complex

We sought to understand how the occurrence of ISI violations depends on underlying cluster FDR and how other underlying characteristics of neuronal activity might affect this relationship. To this end, we focused on three variables likely to have a meaningful effect on the occurrence of ISI violations: neuronal firing frequency (Fig. 1A), temporal correlation of activity among the recorded population of neurons (Fig. 1B), and the number of contaminant neurons (Fig. 1C).

To determine how these variables affect the relationship between ISI_v and FDR, we used Monte Carlo simulations of neural spiking to examine how the relationship between ISI_v and FDR might change as a consequence of varying each parameter in isolation. Total cluster firing frequency was found to have a dramatic effect on ISI violation production, since overall firing frequency played a large role in determining the likelihood that any given FP spike would produce an ISI violation. Critically, clusters with lower firing frequencies and clusters with higher firing frequencies could both present with the same ISI_v, even when they had markedly different underlying FDRs (5–50%; Fig. 2A). These results indicated a nonlinear relationship between ISI violation production and firing frequency not accounted for by simply dividing the number of ISI violations by the number of spikes (ISI_v). The temporal overlap between TPs and FPs was also found to strongly modulate ISI violation production, although not as strongly as cluster firing frequency (Fig. 2B). Variable
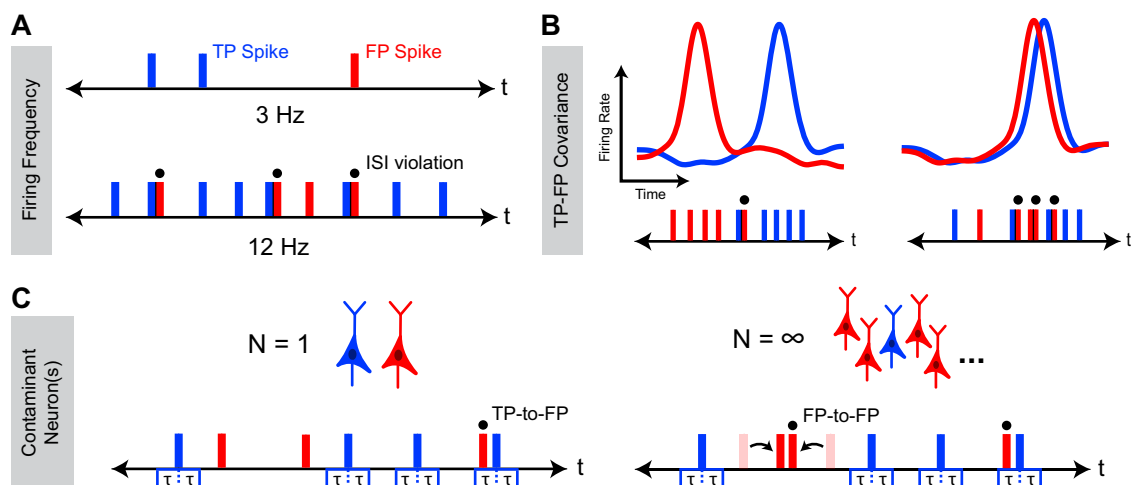

Figure 1. Factors affecting the relationship between ISI_v and FDR. Schematic representation of occurrence of ISI violations for a cluster with varying firing frequencies (**A**) TP–FP covariance (**B**) and numbers of contaminant neuron(s) (**C**). In all cases, underlying FDR between the two cases is the same, while observed ISI_v varies as a consequence of changes in these characteristics of neuronal activity. Blue corresponds to TP spikes, and red corresponds to FP spikes. Overhead dots indicate observed ISI violations. τ represents the neuronal refractory period. ISI violations can occur between TPs and FPs (TP-to-FP) or, if multiple contaminant neurons are present, FPs and other FPs (FP-to-FP).
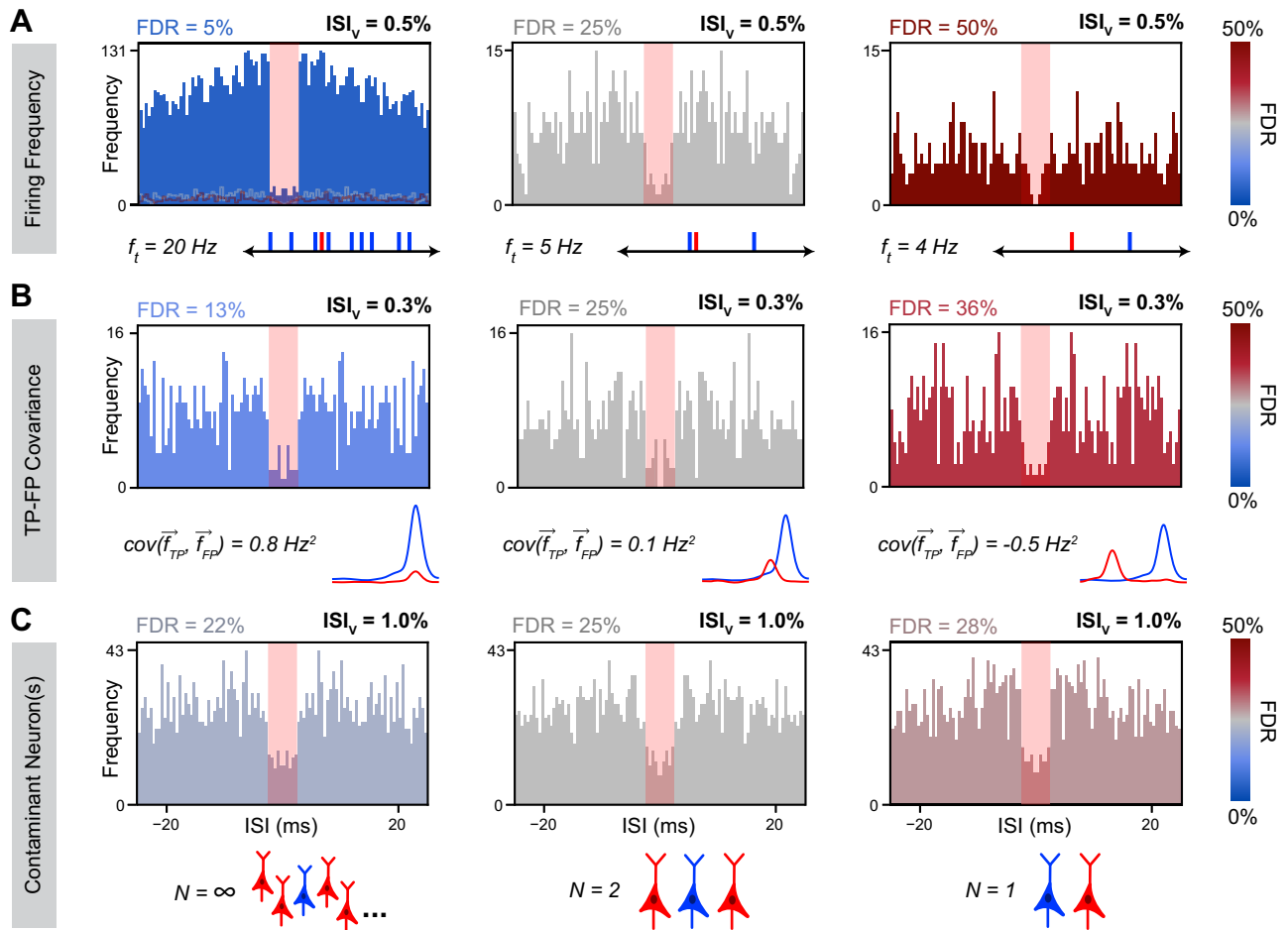
Figure 2. Simulated ISI distributions with variable underlying neuronal characteristics. Representative ISI histograms with all factors kept constant except total cluster firing frequency (**A**), TP–FP covariance (**B**) and contaminant neuron count (**C**). ISIs within the red rectangle are shorter than the refractory period and correspond to ISI violations. ISI_v was identical across conditions in each row. In **A**, ISI violations may appear to constitute a smaller proportion of total ISIs in higher firing frequency conditions, but this is only because these conditions had lower average ISIs and all conditions were plotted within the same 0–25 ms ISI domain. Simulated firing frequencies were as indicated (**A**), 2 Hz (**B**), and 10 Hz (**C**). Simulated firing was either homogeneous (**A**, **C**) or inhomogeneous (**B**). In **A**, outlines of the histograms in the right two panels are overlaid to scale with high transparency on the left panel. Blue corresponds to TP spikes, and red corresponds to FP spikes. FP firing frequency traces (**B**) are schematics only. The case of "infinite" contaminant neurons was simulated using a single contaminant neuron with no refractory period.

TP–FP covariance (0.8 to −0.5 Hz^2) altered the probability of any given FP leading to an ISI violation, resulting in three clusters with substantially different FDRs (13–36%) presenting with the same ISI_v and firing frequency. Lastly, the number of contaminant neurons was also found to modulate ISI violation production, although not as meaningfully as other variables (Fig. 2C). Greater numbers of contaminant neurons increased the odds of ISI violations between pairs of FPs, meaning the same observed ISI_v was associated with a slightly lower FDR in cases of multiple contaminant neurons versus just one contaminant neuron. The dependence of this phenomenon on FP-to-FP violations (Fig. 1C) means its effects only became meaningful at high FDRs (>0.25) and total firing frequencies (>10 Hz).

Analytical model of the relationship between ISI_v and FDR

We next sought to derive an analytical model describing the dependence of underlying FDR on observed ISI_v that incorporates each of these variables. To that end, we first considered a simplified case in which two neurons are each firing homogeneously or at a constant frequency and spikes from both neurons are being assigned to the same cluster. Each TP spike produces a double-sided "violation window" in time. If an FP spike occurs within that window, an ISI violation is observed (Eq. 2; Fig. 1C) as follows:

$$f_v = f_{FP}(2\tau f_{TP}). \tag{2}$$

Here, we represent the number of ISI violations observed per second as f_v, the neuronal absolute refractory period as τ, and the frequencies with which the two neurons produce TP and FP spikes as f_{TP} (TP frequency) and f_{FP} (FP frequency).

Note that f_{TP} and f_{FP} may not necessarily represent the total firing frequencies of the neurons producing the TP and FP spikes, only the frequencies at which spikes from those neurons are assigned to a sorted cluster. These can be one and the same, e.g., for a neuron contributing TPs with no FNs. Also note that f_{TP} and f_{FP} cannot be measured experimentally, instead representing unseen parameters of spike train generation. By making substitutions based on a few simple relationships (Eqs. 3–5), Equation 2 can be solved using the quadratic formula, defining FDR as a function of the following experimentally observable variables: ISI_v, τ, and f_t—the total observed firing frequency of the cluster (Eq. 6; see Extended Data for intermediate steps) as follows:

$$ISI_v = \frac{f_v}{f_t}, \tag{3}$$

$$FDR = \frac{f_{FP}}{f_t}, \tag{4}$$

$$f_t = f_{TP} + f_{FP}, \tag{5}$$

$$FDR = \frac{1}{2}\left(1 - \sqrt{1 - \frac{2ISI_v}{\tau f_t}}\right). \tag{6}$$

The larger root is ignored, i.e., the term under the square root is subtracted and not added, because the neuron with the most spikes in the cluster is de facto considered the TP-contributing neuron. Some spike-sorting algorithms make use of a censor period, whereby spikes detected within a certain minimum distance, τ_c, of another spike are ignored. In such cases, the size of the violation window produced by each TP spike is shortened by this censor period producing a new effective refractory period: $\tau_e = \tau - \tau_c$. Implementation of τ_e produces Equation 7, which is equivalent to a rearranged form of the equation derived in Hill et al. (2011) as follows:

$$f_v = f_{FP}(2\tau_e f_{TP}). \tag{7}$$

At the other extreme, consider a situation where the spikes comprising a cluster are generated by an infinite number of neurons. In reality, there can never be an infinite number of contaminant neurons, but this term is a convenient shorthand for describing the limiting case where there as many contaminant neurons as FPs. In this limiting case, any FP can produce an ISI violation with any other FP, necessitating the addition of a second term wherein FP spikes now produce double-sided violation windows as well. This term is scaled by a factor of one-half to prevent double counting of FPs producing ISI violations with one another. Implementation of this term produces Equation 8, which is equivalent to a rearranged form of the equation derived in Llobet et al. (2022). This equation can be solved for FDR using substitutions and the quadratic formula as previously as the following:

$$f_v = f_{FP}(2\tau_e f_{TP}) + \frac{1}{2}f_{FP}(2\tau_e f_{FP}). \tag{8}$$

For an unspecified number of contaminant neurons N, an additional scaling factor of $(N-1)/N$ can be added to the FP–FP ISI violations term (Eq. 9). This factor can be interpreted as the fraction of all FPs available for any given contaminant neuron's spikes to produce ISI violations with, e.g., 1/2 for $N = 2$ and 2/3 for $N = 3$. This equation, like previous iterations, can be rearranged to solve for FDR with an additional dependence added on N as follows:

$$f_v = f_{FP}(2\tau_e f_{TP}) + \frac{1}{2}\left(\frac{N-1}{N}\right)f_{FP}(2\tau_e f_{FP}). \tag{9}$$

In the more biologically relevant case of inhomogeneous firing, or neural spiking frequencies that vary over time, f_{TP}, f_{FP}, f_t, and f_v can all be considered not as constants but as functions of time (i.e., vector quantities). While this spiking nonstationarity must be taken into account, a time-varying estimate of FDR would be a needless level of granularity and also highly inaccurate given the stochastic nature of neuronal spiking and ISI violations, so of primary interest is a time-averaged estimate of the relationship between violation frequency ($\bar{f}_v$) and underlying variables. In this case, the frequency of violations depends not on the product of the average values of f_{TP} and f_{FP} but on the expected value of their element-wise product, $\mathbb{E}[\vec{f}_{TP} \vec{f}_{FP}]$, as follows:

$$\bar{f}_v = 2\tau_e\mathbb{E}[\vec{f}_{TP} \vec{f}_{FP}] + \frac{1}{2}\left(\frac{N-1}{N}\right)2\tau_e\mathbb{E}[\vec{f}_{FP} \vec{f}_{FP}]. \tag{10}$$

For two vectors representing firing frequency over time $\vec{f}_{TP}$ and $\vec{f}_{FP}$ of length n elements, this expected value can be calculated as follows:

$$\mathbb{E}[\vec{f}_{TP}\,\vec{f}_{FP}] = \frac{\vec{f}_{TP} \cdot \vec{f}_{FP}}{n}. \tag{11}$$

Equation 10 can then be solved for $|\vec{f}_{FP}|$, the vector magnitude of $\vec{f}_{FP}$, using the quadratic formula (Eq. 12). Unit vector $\hat{f}_{FP}$ can subsequently be scaled by $|\vec{f}_{FP}|$, averaged, and finally divided by the mean total firing frequency of the cluster to obtain a time-averaged estimate of FDR (Eq. 13) as follows:

$$|\vec{f}_{FP}| = \frac{N}{N+1}\left(D|\vec{f}_t| - \sqrt{D^2|\vec{f}_t|^2 - \left(\frac{N+1}{N}\right)\left(\frac{\bar{f}_t \mathrm{ISI_v} n}{\tau_e}\right)}\right), \tag{12}$$

$$\mathrm{FDR} = \left(\frac{1}{\bar{f}_t}\right)\frac{1}{n}\sum\left(|\vec{f}_{FP}|\hat{f}_{FP}\right) = \frac{\bar{f}_{FP}}{\bar{f}_t}. \tag{13}$$

Here, D corresponds to the dot product of the unit vectors representing total cluster spike frequency and cluster FP spike frequency, $D = \hat{f}_t \cdot \hat{f}_{FP}$. This can be thought of more generally as representing the degree to which the time-varying total cluster spike frequency temporally overlaps with the time-varying FP spike frequency. This final equation depends on a number of parameters specific to each cluster to obtain a single FDR estimate: (1) the effective refractory period τ_e, (2) the temporal distribution of activity in the cluster of interest ($\vec{f}_t$) and (3) of other clusters contributing FP spikes ($\hat{f}_{FP}$), (4) the observed $\mathrm{ISI_v}$, and (5) the number of contaminant neurons N. For information on how one can estimate these parameters from experimental data, see Materials and Methods.

Prediction of FDR in silico

To both assess our model's predictive power and more generally illustrate relationships between model variables, we next simulated neural spike trains while varying all relevant parameters across a range of biologically relevant values and then attempted to predict FDR from the observed $\mathrm{ISI_v}$. In this case, parameters like N and $\hat{f}_{FP}$ that may normally have associated uncertainty are known exactly. Simulating long periods of time (up to 28 h of recording time) also enables highly accurate estimates of $\mathrm{ISI_v}$.

For both homogeneous and inhomogeneous firing, we found that analytical FDR predictions closely approximated the true underlying FDRs (Fig. 3). For homogeneous firing specifically, observed $\mathrm{ISI_v}$ was found to strongly depend on FDR and total cluster firing frequency, as expected (Fig. 3A). A linear dependence was observed of $\mathrm{ISI_v}$ on firing frequency at fixed FDR and contaminant neuron counts, despite $\mathrm{ISI_v}$ often being assumed to already have normalized for cluster firing frequency. Furthermore, FDR was found to scale quadratically with increasing $\mathrm{ISI_v}$ at a fixed firing frequency and contaminant neuron count (Fig. 3B). $\mathrm{ISI_v}$ values of 0.1–1% represent typical thresholds in literature for considering a cluster well isolated (Jadhav et al., 2009; Roy and Wang, 2012; Chandrasekaran et al., 2017; Wright et al., 2021; Boucher et al., 2023; Zhao et al., 2023). Yet, our results indicate that FDRs associated with these $\mathrm{ISI_v}$ values vary considerably with the firing frequency of the cluster in question (Fig. 3A, B). As an illustration, an $\mathrm{ISI_v}$ of 0.5% reflected a desirable 5% FDR for a cluster firing at 20 Hz or a much higher 50% FDR for one firing at 3 Hz. In general, contaminant neuron count was of limited consequence unless the cluster in question had both a high firing frequency and high FDR, making FP–FP violations frequent enough to meaningfully affect overall ISI violation incidence.

When the activity of clusters was inhomogeneous in time, errors in FDR predicted with the homogeneous model (Eq. 9) scaled linearly with the temporal covariance of TPs and FPs (Fig. 3C). Positive covariance increased the $\mathrm{ISI_v}$ at the same FDR, resulting in overestimation of FDR, while negative covariance conversely decreased ISI violations, resulting in underestimation of FDR. When inhomogeneous firing was appropriately taken into account (Eq. 10), predicted FDR closely approximated true FDR regardless of the covariances in neuronal firing. When the activity of clusters contributing TP spikes and FP spikes varied independently in time (inhomogeneous but zero covariance), the probability of TP–FP violations was unchanged compared with homogeneous spiking. Therefore, under these conditions, the homogeneous and inhomogeneous predictions agree even if the generation of TPs and FPs individually may not necessarily be homogeneous.

We next attempted to predict FDR from $\mathrm{ISI_v}$ (Eqs. 12, 13) in a population of clusters simulated inhomogeneously across a broad, randomized parameter space. Regardless of the precise combination of parameters for any given cluster, predicted FDR and true FDR agreed with low root mean square error (RMSE; Fig. 3D; RMSE, 0.02). This prediction accuracy also generalized to simulations of non-Poisson point processes across a range of possible ISI coefficients of variation (Extended Data Fig. 3-1; RMSE, 0.01–0.02).

Prediction of FDR under realistic conditions

Our model performed well when $\mathrm{ISI_v}$ and other parameters were known exactly, although benchmarking simulations indicated that predictions of FDR are sensitive to small changes in $\mathrm{ISI_v}$ (Fig. 3), particularly at lower firing frequencies
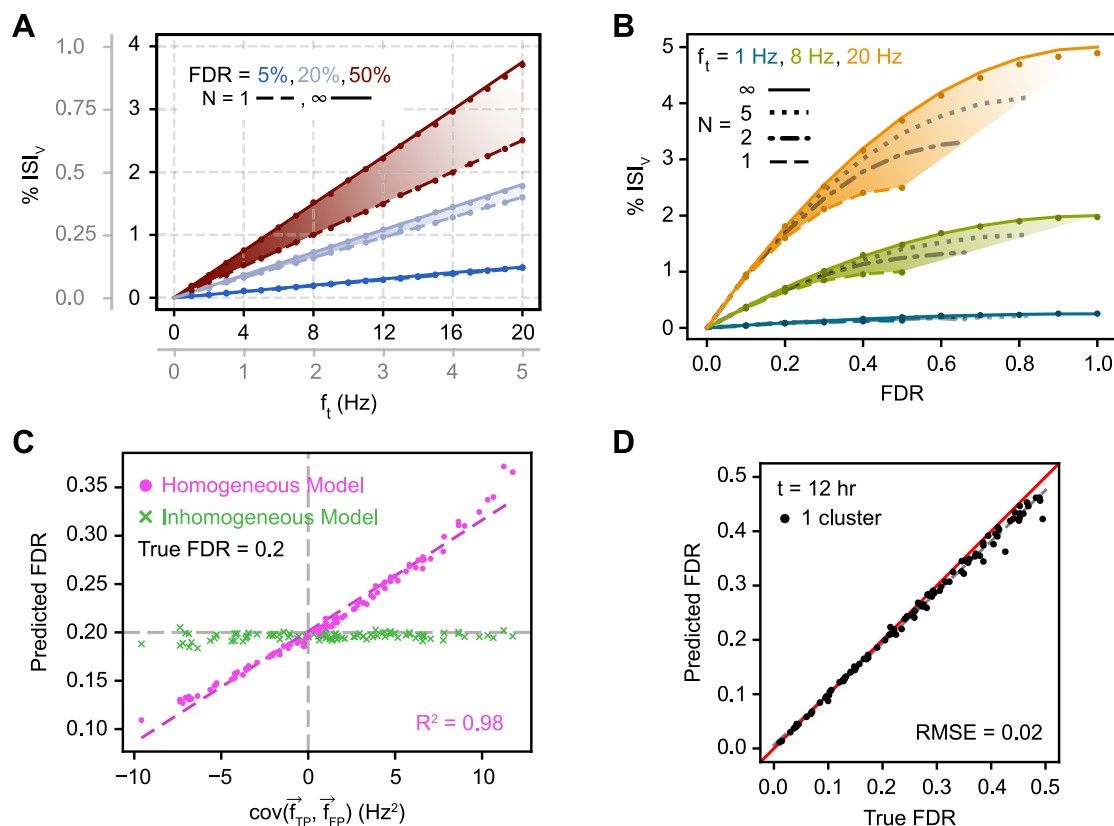
Figure 3. Relationship between single-cluster FDR and observed ISI_v. **A**, Dependence of ISI_v on total firing frequency given varying FDRs and contaminant neuron counts. Lines correspond to analytical predictions; dots correspond to simulation results. Plotted data apply to both primary and gray axes; the latter provide an inset for the former. **B**, Dependence of ISI_v on FDR given varying firing frequencies and contaminant neuron counts. Conventions as in **A**. **C**, Prediction of FDR from observed ISI_v with temporally inhomogeneous firing frequencies using either the homogeneous model (Eq. 9) or the inhomogeneous model (Eq. 10). **D**, Prediction of FDR from observed ISI_v for 100 total clusters simulated across a range of physiologically relevant underlying neuronal characteristics. Total firing frequency was varied between 4 and 20 Hz, N was varied between 1 and ∞, and $\hat{f}_{FP}$ was obtained by averaging across other clusters (see Materials and Methods for more details). The red line is the unity line, or perfect concurrence between predicted and true FDR; dashed gray line is the line of best fit. RMSE calculated with respect to the unity line. Clusters shown in **D** simulated using Poisson point processes. See Extended Data Figure 3-1 for validation using non-Poisson point processes. Simulated firing was either homogeneous (**A**, **B**) or inhomogeneous (**C**, **D**). The model used for FDR predictions was either homogeneous (**A**, **B**) or inhomogeneous (**D**), or both were used (**C**).

(Fig. 3*B*). We next wanted to determine whether FDR could be accurately predicted from recordings of finite duration with noisy ISI_v estimates and when exact values of N and $\hat{f}_{FP}$ are unknown. Simulating spiking for finite durations, we found that observed ISI_v values were normally distributed around their true values (Fig. 4*A*), with increasing recording time decreasing the variance of this distribution, as expected.

To assess the effect of noisy ISI_v estimates on FDR predictions, we again simulated spiking while simultaneously varying all previously described parameters, but this time recording duration was restricted to 10 min (Fig. 4*B,C*). In this case, we found that FDR predictions for individual clusters were substantially less accurate (RMSE, 0.13), although they did not deviate systematically from true FDRs (Fig. 4*B*). Even with this highly restricted recording time, FDR population statistics could be estimated across a set of clusters with high fidelity so long as a sufficient number of clusters were sampled (Fig. 4*C*). For example, median FDR and mean FDR could be predicted with RMSEs of 0.03 and 0.02, respectively, across 1,000 clusters. Predicted median FDRs were slightly overestimated at high true median FDR (>0.25).

We next sought to determine the duration of neural recordings necessary to obtain accurate estimates of ISI_v and, thus, FDR in single clusters. To accomplish this, we simulated clusters with various firing frequencies and FDRs and determined the recording time sufficient to produce a CV of the predicted FDR of 20% (Fig. 4*D*; e.g., 50 ± 10% or 5 ± 1%). Surprisingly, we found that clusters with firing frequencies of 1–2 Hz required observations across hundreds of minutes to produce accurate estimates of single-cluster FDR, an infeasible recording time in many common experimental paradigms. For clusters with firing frequencies >5 Hz, FDRs between 5 and 30% could be estimated accurately using tens of minutes of spiking data.

We used our model to estimate the FDRs of clusters contained within 12 publicly available datasets that included spike-sorted electrophysiology recordings (Table 2). For all datasets, the inhomogeneous model (Eq. 10) was used,
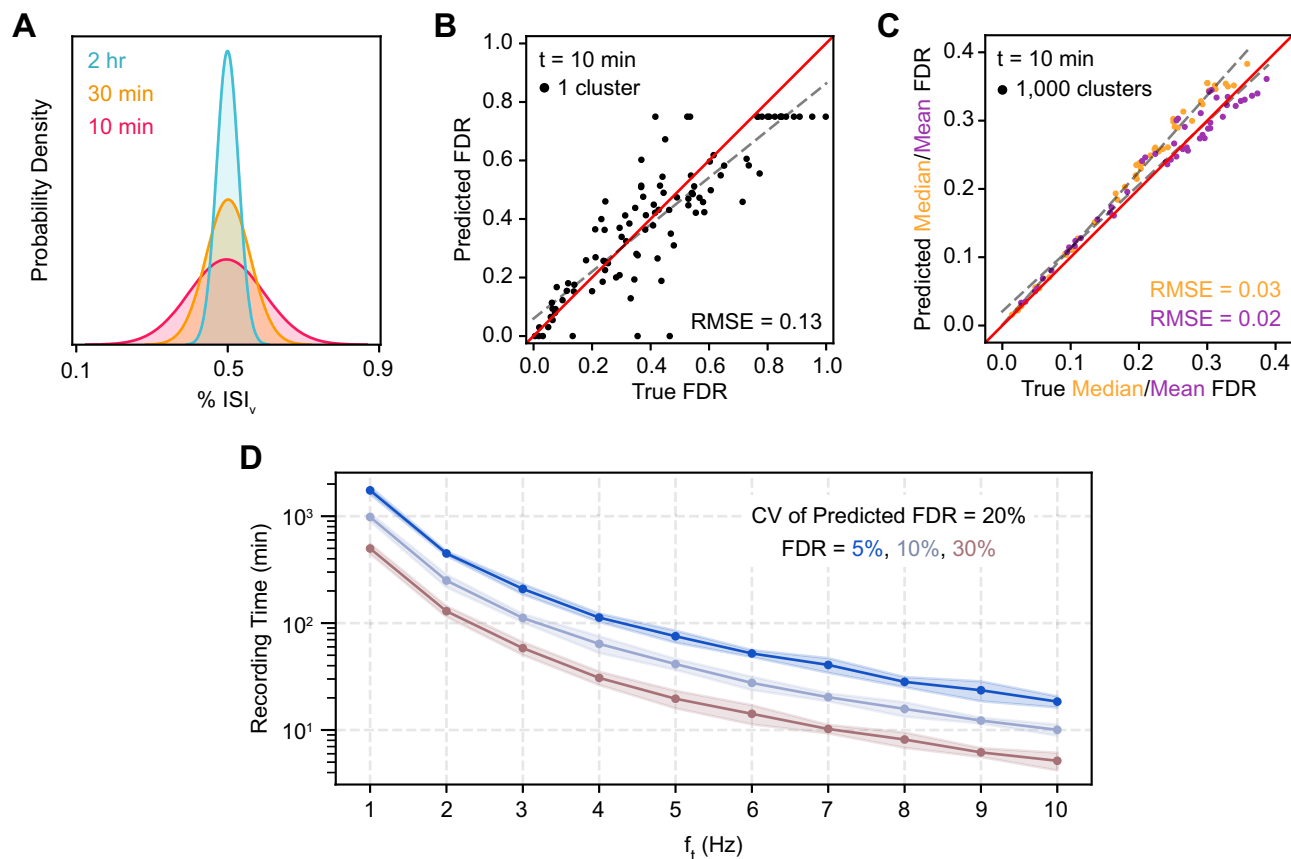
Figure 4. Prediction of FDR for single clusters and populations. **A**, Probability density functions of observed % ISI$_v$ for a prototypical 8 Hz, 15% FDR cluster recorded for varying time lengths. These parameters produce a true % ISI$_v$ of 0.5%, a value that has been used as a threshold for considering clusters well isolated (Jadhav et al., 2009; Guo et al., 2014; Chandrasekaran et al., 2017). **B**, Prediction of FDR in single clusters recorded for 10 min. All parameters varied simultaneously across a range of physiologically relevant underlying neuronal characteristics. Total firing frequency was varied between 1 and 10 Hz, N was varied between 1 and ∞, and $\hat{f}_{FP}$ was obtained by averaging across other clusters (see Materials and Methods for more details). Ceiling effect of predicted FDR at 0.75 due to theoretical limit of FDR when predicting with unknown N (see Materials and Methods). The red line is the unity line or perfect concurrence between predicted and true FDR; the dashed gray line is the line of best fit. RMSE calculated with respect to the unity line. **C**, Prediction of median and mean FDR in 1,000-cluster populations. Cluster FDRs are Cauchy-distributed around the population mean. Total firing frequency was varied between 4 and 16 Hz; otherwise parameters varied simultaneously and conventions as in **B**. **D**, Minimum recording time required for predictions of FDR in a single cluster to have a CV of 20%.

with the exception of Stringer et al. (2019) and Juavinett et al. (2019). Animal behavior in these datasets was not trial-based, and so it was not straightforward to accurately estimate time-varying firing frequencies, necessitating the use of the homogeneous model (Eq. 9; see Materials and Methods). Given limitations associated with single-cluster FDR predictions (Fig. 4B), estimated median and mean FDRs were reported across all clusters present in each dataset. An average median FDR of 12.9 ± 13.5% (SD) was observed along with an average mean FDR of 24.1 ± 9.2% (SD). No obvious correlations between cluster count, spike-sorting methodology, or recording technology and dataset FDR were observed. With the exception of Juavinett et al. (2019), median estimated FDRs were consistently lower than mean estimated FDRs. This implies that cluster FDR distributions in recorded electrophysiology datasets tend to be right-skewed, composed of a large proportion of clusters possessing FDRs closer to 0 as well as a broadly distributed complement of more contaminated clusters, some of which potentially reaching FDRs well above 0.5.

Importantly, no attempts were made to curate clusters included in analysis from each dataset; all available sessions and clusters were assessed, with the exception of clusters labeled "multi" or "MUA" in datasets with cluster quality annotations. It's possible that some datasets were precurated, with clusters discarded according to some exclusion criteria prior to being uploaded, while others were shared uncurated. Depending on the nature of the scientific question being addressed, better cluster isolation is undoubtedly a larger priority for some datasets than others. For these reasons and due to uncertainty in assessing censor period in each dataset, we emphasize that these findings should serve generally as an overall survey of the range of expected contamination levels in datasets produced using widely used methods rather than as a commentary on individual datasets or spike-sorting methodologies.

Table 2. Median and mean FDR of publicly available spike-sorted electrophysiology datasets

Authors	Median FDR (±SE)	Mean FDR (±SE)	# Clusters	τ_c (ms)	Probe	Sorter
Xu et al. (2022)	3.1% (0.5)	14.3% (0.4)	3,046	0	H2/H3 Cam. Neuro.	Kilosort
Economo et al. (2018)	3.1% (0.6)	12.5% (0.5)	1,988	0.25	H2 Cam. Neuro.	JRClust
Steinmetz et al. (2019)[a]	4.0% (0.2)	25.1% (0.2)	33,997	0	Neuropixels 1.0	Kilosort
Gao et al. (2018)	4.4% (0.7)	18.7% (0.6)	1,923	0.5	H2/A4 Cam. Neuro./NeuroNexus	UltraMegaSort2000
Li et al. (2016)	4.5% (0.8)	19.7% (0.7)	1,543	0.85	A4 NeuroNexus	UltraMegaSort2000
Inagaki et al. (2022)	6.2% (0.3)	17.7% (0.3)	7,968	0.25	HH-2/Neuropixels 2.0 Janelia/N.A.	JRClust/Kilosort
Guo et al. (2017)	6.5% (0.7)	18.8% (0.6)	1,936	0.85	A4/A2 NeuroNexus	UltraMegaSort2000
Finkelstein et al. (2021)[a]	10.5% (0.5)	20.1% (0.4)	3,385	0.25	H2 Cam. Neuro.	JRClust
Sylwestrak et al. (2022)	12.8% (0.4)	29.1% (0.3)	10,548	0.25	Neuropixels 2.0	Kilosort
Stringer et al. (2019)[b]	21.9% (0.5)	33.1% (0.4)	6,446	0	Neuropixels 1.0	Kilosort
Chinta and Pluta (2023)	28.2% (1.3)	36.2% (1.0)	991	0	NeuroNexus	Kilosort
Juavinett et al. (2019)[a,b]	50.0% (0.8)	44.3% (0.7)	2,018	0	Neuropixels 1.0	Kilosort

[a]Clusters labeled "multi" or "MUA" excluded.
[b]Calculated using homogeneous model (Eq. 9).

Discussion

ISI violations are the most commonly employed metric of accuracy in spike sorting, serving as an indication of the FDR—the rate at which spikes are erroneously assigned to the wrong cluster. Here, we used Monte Carlo simulations to demonstrate that the ISI_v is related to FDR through a complex relationship that depends on many factors, including (1) the neuronal firing frequency, (2) the temporal correlation in activity between neurons contributing to a cluster, and (3) the number of neurons contributing spikes to a cluster—in that order of descending importance. We derived an analytical model that can be used to predict FDR from ISI_v that incorporates these factors and determine the accuracy with which FDR can be inferred during finite-length recordings at the level of single clusters and datasets. Finally, we used this model to assess the FDR of clusters contained in publicly available spike-sorted electrophysiology datasets to provide bounds on the accuracy that can be reasonably expected by experimenters. Our study makes four central contributions.

First, we derive an analytical model that can be used to estimate FDR from ISI_v accurately across a broad parameter space.

Second, we explicitly demonstrate that FDR is not linearly related to ISI_v but depends critically on the total cluster spike frequency. While this dependence can be inferred from previous work (Hill et al., 2011; Llobet et al., 2022), our results underscore the inappropriateness of using ISI_v as an inclusion criterion for single clusters—which is still a common practice in many studies using spike-sorted data. Across a common range of firing frequencies (~2 to 12 Hz), clusters with the same ISI_v can be associated with both low (~5%) and very high (~50%) FDRs (Figs. 2A, 3A).

Third, we find that estimates of FDR at the single-cluster level are noisy due to the stochasticity of ISI violations as well as uncertainty in cluster-specific parameters (Fig. 4B)—but estimates at the population level are highly robust (Fig. 4C). As a point of reference, our results suggest that ISI_v can be estimated accurately enough to predict single-cluster FDRs within 20% of their true values in 1 h recordings only when firing frequencies are greater than ~5 Hz. Even for clusters that meet these requirements, however, experimental uncertainty in parameters like FP–TP covariance and contaminant neuron counts make single-cluster FDRs difficult to predict with high confidence. Alternatively, population-level statistics of FDR, obtained by averaging across all the clusters in a dataset, can be accurately predicted with recording durations as low as 10 min.

Finally, we predict FDR on the basis of ISI violations in publicly available datasets for the first time. FDR population statistics covered a wide range (median, 3.1–50.0%; mean, 12.5–44.3%) and could be estimated with low standard error (SE median, 0.2–1.3%; SE mean, 0.2–1.0%). Datasets with low FDR were not associated with any obvious external features of data collection, spike-sorting methodology, or recording technology, and the variance in FDRs across datasets is likely a function of the variable importance of low FDR for the scientific goals of individual studies.

Monte Carlo simulations for validating prediction of FDR

Given the absence of high cluster count spike-sorted extracellular neural recordings with associated per-neuron ground truth patch–clamp data, Monte Carlo simulation presents itself as an attractive tool for studying the theoretical mechanisms by which a given level of contamination in a sorted cluster translates into observed ISI violations. Neural spiking was simulated in this work as a collection of independent Poisson processes, a common assumption that has been validated across a number of organisms, brain regions, and behavioral contexts (Werner and Mountcastle, 1965; Tolhurst et al., 1981; Shinomoto and Tsubo, 2001; Abbott and Dayan, 2005; Roxin et al., 2008). Furthermore, the equations derived in this work were equally accurate when applied to simulations of non-Poisson point processes across a range of ISI coefficients of variation, implying that they likely generalize to any renewal process that produces independent and identically distributed ISIs with a physiological CV (Extended Data Fig. 3-1). Beyond this foundation, only aspects of neuronal firing and spike sorting thought to potentially be relevant to ISI violation production were modeled, namely, varying firing frequency amplitudes, temporally inhomogeneous firing, and differing contaminant neuron counts. If an additional

aspect not accounted for in this description plays a significant role in determining the relationship between ISI_v and FDR, then in silico validation may not reflect true congruence between analytical prediction and reality.

Best practices for using ISI violations in spike sorting

When attempting to determine the success of spike-sorting operations post hoc, ISI violation fraction is frequently used as a per-cluster inclusion criterion. However, unless a cluster is recorded for a long-enough time period given its firing frequency and true FDR and difficult-to-estimate cluster–specific parameters are known, it can be difficult to predict FDR using ISI violations at the single-cluster level with high confidence. The use of ISI violation fractions in this way can easily result in situations where highly contaminated clusters are erroneously kept, while less contaminated clusters are discarded. We posit that the most straightforward and robust use case for ISI_v is as a tool for predicting population-level statistics of FDR when coupled with a sound theoretical understanding of how cluster contamination translates into ISI violations (Eqs. 10, 12, 13). Investigators can obtain an accurate estimate of median and mean cluster FDR across a session or dataset and then decide whether they are satisfied with these levels of cross-contamination or if additional effort to improve cluster isolation is needed.

When curating spike-sorted data, it is critical that both algorithmic and manual sorters do not specifically remove individual spikes that generate ISI violations. Typically, only a small fraction of contaminant spikes produce ISI violations; targeted removal of spikes producing ISI violations can reduce ISI_v substantially without meaningfully reducing the FDR, thus producing clusters that seem well isolated based on their ISI_v, even when they are not. In effect, this practice does not accomplish anything except eliminating the predictive power of ISI_v for underlying FDR.

Current state of spike sorting predicted using ISI violations

This work estimates an average median FDR of ~13% and an average mean FDR of ~24% in publicly available electrophysiology datasets. The lower median FDR compared with mean FDR across virtually all datasets examined indicates right-skewed FDR distributions. This likely arises as a consequence of FDR having a theoretical floor of 0, with most datasets having many cluster FDRs close to this floor. It may also be a consequence of most spike-sorting datasets being composed of two distinct types of clusters: relatively easier-to-sort clusters whose FDRs typically fall close to 0 and relatively harder-to-sort clusters whose FDRs are likely to fall more broadly over the theoretical range of FDR (0–1).

In the absence of any clear rationale for variance in FDR as predicted using ISI_v among publicly available spike-sorted datasets (Table 2), degree and execution of manual curation present themselves as promising explanatory candidates. While modern spike-sorting algorithms serve as an excellent basis for sorting vast quantities of electrophysiology data, many investigators still manually merge, split, and discard algorithmically obtained output clusters to further improve cluster isolation. Time, effort, and skill applied to manual curation are difficult to quantify and therefore unlikely to be reported in literature, although such differences are likely to have a material effect on the final quality of cluster isolation. The median and mean FDRs of the datasets examined here as well as the tendency toward right-skewed FDR distributions support the idea that all datasets are composed of both well and poorly isolated clusters.

Necessity of high-quality spike sorting

It has been posited that well-sorted clusters are not a necessity for many types of neural data analyses, particularly those concerned with studying population dynamics (Christie et al., 2014; Trautmann et al., 2019). In some applications, however, well-isolated clusters remain a critical precondition for answering relevant neuroscientific questions. Characterizing the responsivity of specific cell types that have been identified on the basis of genetic expression, projection target, waveform shape, or activity in vivo represents an expansive line of inquiry wherein high-quality cluster isolation is key (Estebanez et al., 2017; Deubner et al., 2019; E. K. Lee et al., 2021; Ding et al., 2022; Takatoh et al., 2022). Ultimately, the level of cluster isolation necessary for a given study is highly dependent upon the biological questions of interest. The work herein aims to clarify the relationship between ISI violations and cluster contamination, as well as provide a tool by which overall spike-sorting quality can be quickly assessed with a direct, interpretable, and accurate metric, thereby streamlining assessments of sorting performance and increasing confidence that desired cluster isolation levels have been reached.

References

Abbott LF, Dayan P (2005) *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Barnett AH, Magland JF, Greengard LF (2016) Validation of neural spike sorting algorithms without ground-truth information. J Neurosci Methods 264:65–77.

Bestel R, Daus AW, Thielemann C (2012) A novel automated spike sorting algorithm with adaptable feature extraction. J Neurosci Methods 211:168–178.

Boucher PO, Wang T, Carceroni L, Kane G, Shenoy KV, Chandrasekaran C (2023) Initial conditions combine with sensory evidence to induce decision-related dynamics in premotor cortex. Nat Commun 14:6510.

Buccino AP, Hurwitz CL, Garcia S, Magland J, Siegle JH, Hurwitz R, Hennig MH (2020) Spikeinterface, a unified framework for spike sorting. Elife 9:e61834.

Chandrasekaran C, Peixoto D, Newsome WT, Shenoy KV (2017) Laminar differences in decision-related neural activity in dorsal premotor cortex. Nat Commun 8:614.

Chaure FJ, Rey HG, Quian Quiroga R (2018) A novel and fully automatic spike-sorting implementation with variable number of features. J Neurophysiol 120:1859–1871.

Chinta S, Pluta SR (2023) Neural mechanisms for the localization of unexpected external motion. Nat Commun 14:6112.

Christie BP, Tat DM, Irwin ZT, Gilja V, Nuyujukian P, Foster JD, Ryu SI, Shenoy KV, Thompson DE, Chestek CA (2014) Comparison of spike sorting and thresholding of voltage waveforms for intra-cortical brain–machine interface performance. J Neural Eng 12:016009.

Chung JE, Magland JF, Barnett AH, Tolosa VM, Tooker AC, Lee KY, Shah KG, Felix SH, Frank LM, Greengard LF (2017) A fully auto-mated approach to spike sorting. Neuron 95:1381–1394.e6.

Deubner J, Coulon P, Diester I (2019) Optogenetic approaches to study the mammalian brain. Curr Opin Struct Biol 57:157–163.

Ding L, Balsamo G, Chen H, Blanco-Hernandez E, Zouridis IS, Naumann R, Preston-Ferrer P, Burgalossi A (2022) Juxtacellular opto-tagging of hippocampal CA1 neurons in freely moving mice. Elife 11:e71720.

Economo MN, et al. (2018) Distinct descending motor cortex pathways and their roles in movement. Nature 563:79–84.

Estebanez L, Hoffmann D, Voigt BC, Poulet JFA (2017) Parvalbumin-expressing GABAergic neurons in primary motor cortex signal reaching. Cell Rep 20:308–318.

Finkelstein A, Fontolan L, Economo MN, Li N, Romani S, Svoboda K (2021) Attractor dynamics gate cortical information flow during decision-making. Nat Neurosci 24:843–850.

Gao Z, Davis C, Thomas AM, Economo MN, Abrego AM, Svoboda K, De Zeeuw CI, Li N (2018) A cortico-cerebellar loop for motor plan-ning. Nature 563:113–116.

Gibson S, Judy JW, Marković D (2012) Spike sorting: the first step in decoding the brain: the first step in decoding the brain. IEEE Signal Process Mag 29:124–143.

Guo ZV, Inagaki HK, Daie K, Druckmann S, Gerfen CR, Svoboda K (2017) Maintenance of persistent activity in a frontal thalamocorti-cal loop. Nature 545:181–186.

Guo ZV, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, Feng G, Svoboda K (2014) Flow of cortical activity underlying a tactile decision in mice. Neuron 81:179–194.

Harris KD, Hirase H, Leinekugel X, Henze DA, Buzsáki G (2001) Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. Neuron 32:141–149.

Hill DN, Mehta SB, Kleinfeld D (2011) Quality metrics to accompany spike sorting of extracellular signals. J Neurosci 31:8699–8705.

Inagaki HK, Chen S, Ridder MC, Sah P, Li N, Yang Z, Hasanbegovic H, Gao Z, Gerfen CR, Svoboda K (2022) A midbrain-thalamus-cortex circuit reorganizes cortical dynamics to initiate movement. Cell 185:1065–1081.e23.

Jadhav SP, Wolfe J, Feldman DE (2009) Sparse temporal coding of elementary tactile features during active whisker sensation. Nat Neurosci 12:792–800.

Juavinett AL, Bekheet G, Churchland AK (2019) Chronically implanted neuropixels probes enable high-yield recordings in freely moving mice. Elife 8:e47188.

Jun JJ, Mitelut C, Lai C, Gratiy SL, Anastassiou CA, Harris TD (2017). Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction (p. 101030). bioRxiv.

Lee EK, Balasubramanian H, Tsolias A, Anakwe SU, Medalla M, Shenoy KV, Chandrasekaran C (2021) Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. Elife 10:e67490.

Lee JH, Carlson DE, Shokri Razaghi H, Yao W, Goetz GA, Hagen E, Batty E, Chichilnisky EJ, Einevoll GT, Paninski L (2017) YASS: yet another spike sorter. Adv Neural Inf Process Syst 30:4002–4012.

Li N, Daie K, Svoboda K, Druckmann S (2016) Robust neuronal dynamics in premotor cortex during motor planning. Nature 532:459–464.

Llobet V, Wyngaard A, Barbour B (2022). Automatic post-processing and merging of multiple spike-sorting analyses with Lussac (p. 2022.02.08.479192). bioRxiv.

Magland J, Jun JJ, Lovero E, Morley AJ, Hurwitz CL, Buccino AP, Garcia S, Barnett AH (2020) Spikeforest, reproducible web-facing ground-truth validation of automated neural spike sorters. Elife 9:e55167.

Neymotin SA, Lytton WW, Olypher AV, Fenton AA (2011) Measuring the quality of neuronal identification in ensemble recordings. J Neurosci 31:16398–16409.

Pachitariu M, Steinmetz NA, Kadir SN, Carandini M, Harris KD (2016) Fast and accurate spike sorting of high-channel count probes with Kilosort. Adv Neural Inf Process Syst 29:4448–4456.

Pouzat C, Mazor O, Laurent G (2002) Using noise signature to optimize spike-sorting and to assess neuronal classification quality. J Neurosci Methods 122:43–57.

Quiroga RQ, Panzeri S (2013) *Principles of neural coding*. Boca Raton, FL: CRC Press.

Rey HG, Pedreira C, Quian Quiroga R (2015) Past, present and future of spike sorting techniques. Brain Res Bull 119:106–117.

Roxin A, Hakim V, Brunel N (2008) The statistics of repeating patterns of cortical activity can be reproduced by a model network of sto-chastic binary neurons. J Neurosci 28:10734–10745.

Roy S, Wang X (2012) Wireless multi-channel single unit recording in freely moving and vocalizing primates. J Neurosci Methods 203:28–40.

Saif-ur-Rehman M, et al. (2021) Spikedeep-classifier: a deep-learning based fully automatic offline spike sorting algorithm. J Neural Eng 18:016009.

Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish AD (2005) Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience 131:1–11.

Shinomoto S, Tsubo Y (2001) Modeling spiking behavior of neurons with time-dependent Poisson processes. Phys Rev E 64:041910.

Steinmetz NA, Zatka-Haas P, Carandini M, Harris KD (2019) Distributed coding of choice, action and engagement across the mouse brain. Nature 576:266–273.

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional, brain-wide activity. Science 364:eaav7893.

Sylwestrak EL, et al. (2022) Cell-type-specific population dynamics of diverse reward computations. Cell 185:3568–3587.e27.

Takatoh J, et al. (2022) The whisking oscillator circuit. Nature 609:560–568.

Todorova S, Sadtler P, Batista A, Chase S, Ventura V (2014) To sort or not to sort: the impact of spike-sorting on neural decoding perfor-mance. J Neural Eng 11:056005.

Tolhurst DJ, Movshon JA, Thompson ID (1981) The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. Exp Brain Res 41:414–419.

Toosi R, Akhaee MA, Dehaqani M-RA (2021) An automatic spike sort-ing algorithm based on adaptive spike detection and a mixture of skew-t distributions. Sci Rep 11:13925.

Trautmann EM, et al. (2019) Accurate estimation of neural population dynamics without spike sorting. Neuron 103:292–308.e4.

van Vreeswijk C (2010) Stochastic models of spike trains. In: *Analysis of parallel spike trains* (Grün S, Rotter S, eds), pp 3–20. New York, NY: Springer US.

Werner G, Mountcastle VB (1965) Neural activity in mechanoreceptive cutaneous afferents: stimulus-response relations, weber functions, and information transmission. J Neurophysiol 28:359–397.

Wright NC, Borden PY, Liew YJ, Bolus MF, Stoy WM, Forest CR, Stanley GB (2021) Rapid cortical adaptation and the role of tha-lamic synchrony during wakefulness. J Neurosci 41:5421–5439.

Xu D, Dong M, Chen Y, Delgado AM, Hughes NC, Zhang L, O'Connor DH (2022) Cortical processing of flexible and context-dependent sensorimotor sequences. Nature 603:464–469.

Yegenoglu A, Denker M, Grün S (2018) Collaborative HPC-enabled workflows on the HBP collaboratory using the elephant framework. INM-ICS Retreat 2018.

Zhao S, Tang X, Tian W, Partarrieu S, Liu R, Shen H, Lee J, Guo S, Lin Z, Liu J (2023) Tracking neural activity from the same cells during the entire adult life of mice. Nat Neurosci 26:696–710.