Interpreting Diagnostic Histories into Infection Time Estimates:  
Framework and Online Tool

Eduard Grebe1#\*, Shelley Facente2\*, Jeremy Bingham1, Reshma Kassanjee3, Andrew Powrie4, Jaryd Gerber4, Gareth Priede4, Trust Chibawara1, Michael Busch5,7, Gary Murphy6, Chris Pilcher7 & Alex Welte1\*

*1South African Centre for Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa; 2Facente Consulting, Richmond, CA, USA; 3Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa; 4Implicit Design, Cape Town, South Africa; 5Blood Systems Research Institute, San Francisco, CA, USA; 6Public Health England, London, United Kingdom; 7University of California San Francisco, San Francisco, CA, USA*

\*These authors contributed equally.

#Email: [eduardgrebe@sun.ac.za](mailto:eduardgrebe@sun.ac.za); Address: SACEMA, Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa.

# Abstract

# Introduction

For pathogenesis studies, diagnostic biomarker evaluation, and surveillance purposes, it is frequently of interest to estimate the infection timing (date of infection or time-since-infection) of study subjects. Ideally, a biomarker signature would provide reasonable direct estimates of an individual’s time-since-infection, but natural inter-subject variability of pathogenesis and disease progression makes this difficult. This work focuses on the use of readily available, qualitative (i.e. positive/negative) test results to estimate dates of infection.

Most simply, nuanced infection dating applies to subjects who produce a negative test and also (usually at a later time) a positive test result, taking into account that no test can detect infection immediately after exposure. Hence, infection can at best be estimated to have occurred during an interval in the past, relative to the date(s) of the test(s).

When a subject obtains ‘discordant results’ – i.e. a negative and a positive test result on the same day, this typically reflects positive results arising on ‘more sensitive’ tests than those on which the negative results were obtained. What we mean here by higher sensitivity is a shorter *‘typical delay between exposure/acquisition and detectability of infection.*’ For high-performing diagnostic tests, such as are normal for HIV and other viral infections like hepatitis, test sensitivity is optimally understood as being a function of time-since-infection, rather than simply the conventional statistical meaning of sensitivity as the probability of correctly identifying a positive case. We further discuss this idea in some detail in a related analysis of the ‘residual risk’ of infection from blood products, that remains despite extensive screening (Welte et al., forthcoming).

For more than 15 years, the only widely referenced explicit infection dating algorithm, that uses test results to estimate time-since-infection has been the ‘Fiebig staging’ system [1], which defines a number of stages of early HIV infection through various ‘standard’ combinations of discordant results using diagnostic tests of different ‘sensitivity.’ For example, Fiebig stage 1 is defined as exhibiting reactivity on a viral load assay, but not (yet) on a p24 antigen assay, and in the seminal 2003 paper was estimated to begin approximately 11 days after infection, with a mean duration of 5.0 days [1]. The particular tests used in these original calculations are no longer in use or commercially available. Others have used newer diagnostic assays to recalibrate the Fiebig stage mean duration estimates or redefine similar stages as an analog to the Fiebig method [2,3], though as testing technology continues to evolve rapidly, attempts to recalibrate these stages become increasingly difficult.

Infection date estimates will usually be well summarised as intervals, the midpoint of which is naturally considered a ‘point estimate’ of the date of infection. These intervals can be understood as (somewhat round shouldered) plateaus, and can be interpreted from a Bayesian point of view as ‘posteriors.’ The intervals can further serve as ‘priors’ in the analysis of additional quantitative markers obtained within a calibrated dynamic range.

Building from the Fiebig staging concept, we developed a new, more nuanced infection dating algorithm to meet the needs of a substantial collaboration (the Consortium for the Evaluation and Performance of HIV Incidence Assays – CEPHIA [4]) in support of the discovery, development and evaluation of recent infection biomarkers [5,6]. The primary CEPHIA activity was to develop various case definitions for ‘recent HIV infection’, with intended applicability mainly to HIV incidence surveillance, rather than individual-level staging [5,7,8], although the latter application was also explored [8]. A key challenge was that, though based on large runs of many specimens under consistent conditions at the Blood Systems Research Institute (BSRI) and Public Health England (PHE), the primary recency test results bring together material from subjects in numerous studies, each of which used different diagnostic algorithms to capture information about the timing of HIV acquisition or seroconversion. To meet this challenge we linked specimens from thousands of study-patient interactions into a coherent and consistent infection dating scheme, in order to estimate the critical properties of recent infection tests for surveillance applications, i.e. Mean Duration of Recent Infection (MDRI) and False Recent Rate (FRR) as previously defined [9].

In order to align diagnostic testing information across multiple sources, one needs a common reference event in a patient history – ideally, the time of infectious exposure. Alas, the timing of infectious exposure is seldom known, even in intensive studies, and studies of diagnostic tests essentially provide *relative* times of test conversion [10–12]. Our dating scheme uses as a reference event the first time that a highly-sensitive viral load assay – with a detection threshold of 1 RNA copy/ml of plasma – could [would] detect an infection. We call this the *Date of* *Detectable Infection* (DDI). Where possible we then endow study subjects with a point estimate of this date, which we call the *Estimated Date of Detectable Infection* (EDDI). The time from infectious exposure to DDI is likely to be variable between individuals, but would be on the order of a few days [13]. We assume that after the viral load reaches 1 RNA copy/ml, viral load increases exponentially during the initial ramp-up phase, with the growth rate having been estimated at 0.35 log10(RNA copies/ml)/day [1].

The present work outlines the analytical framework for this algorithm for infection date estimation using ‘diagnostic testing histories’ [14] and proposes additional nomenclature and strategies to address atypical test result combinations, including test correlation. It introduces a publicly available online HIV infection dating tool that facilitates the use of this algorithm. The tool, detailed in the Appendix, brings together 1) curatorship of HIV test performance data, and 2) infection date estimation functionality. It is readily updatable as test technology evolves, given the simple general architecture of the system and its nature as an open source project.

# Fiebig-like staging

The fundamental feature of the Fiebig staging system [1] is that it identifies a naturally-occurring sequence of *discordant-diagnostic-test* states that indicate early clinical disease progression, and whose approximate duration can be estimated, given a sufficient number of clinically well-characterised specimens. One can simply treat these numbered stages as an arbitrary pathway that infected persons pass through, without necessarily incorporating the underlying discordant test results (such as a positive p24 antigen test and a negative ELISA antibody test) into analysis. However, this makes it substantially less feasible to propose and calibrate variations to the stages, given changes in test technology and usage. As we demonstrate below, it is most robust to interpret any combination of diagnostic test results into their underlying tests, if these tests have been independently benchmarked for diagnostic sensitivity (i.e. a median or mean duration of time from infection to detectability has been estimated).

Interpreted at a population level, a particular test’s sensitivity curve expresses the probability that a specimen from a ‘completely randomly’ chosen subject (i.e. all population members have the same probability of being selected) will produce a positive result, given that the specimen was obtained at a time after either: 1) infectious exposure, or, 2) detectability on some other well-characterised diagnostic test. The key features of a test’s sensitivity curve (represented by the purple curve in figure 1) are that:

* there is effectively no chance of detecting an infection immediately after exposure;
* after some time, the test will almost certainly detect an infection;
* there is a characteristic time range over which this function transitions from close to zero to close to one. This can be summarised as something very much like a mean or median and a standard deviation.

Figure 1: Individual and population-level diagnostic test sensitivity curves



For individuals in the population, the test will perform with some variation from the red population curve. Genotype and other attributes, such as concurrent infections, age, the particular invading quasispecies, post-infection factors, etc. affect the performance of a test for a particular individual; this in principle determines a subject-specific curve, such as one of the blue curves in Figure 1, which capture the probability, as a function of time, that specimens from a particular subject will produce a positive diagnostic result. Because assay results are themselves imperfectly reproducible even on the same individual, these blue curves do not transition step-like from zero to one, but have some finite window of time, though considerably shorter than the dynamic range of the population-level curve, over which they transition from close to zero to close to one.

The *scale* of the blue curves captures something like an individual-level ‘diagnostic delay’, while the *shape* encapsulates 1) the growth rate of the underlying assay response variable *in the region of the positive/negative threshold*, and 2) the measurement noise *in the region of the diagnostic threshold*. The red population-level curve, on the other hand, captures the population-level distribution of the individual diagnostic delays – that is, the distribution of ‘scale parameters’ for the individual (blue) curves. The shape of this curve has no particular relationship to the shapes of the individual sensitivity curves.

In contrast to the usual statistical calculation of ‘sensitivity’ as the proportion of ‘true positive’ specimens that produce a positive result, we propose to summarise the population-level sensitivity of any particular diagnostic test into one or two ‘diagnostic delay’ parameters ( and in Figure 1). By far the most important parameter is an estimate of ‘*median diagnostic delay’*, measured, for present purposes, as time from earliest possible detection by a viral load assay with a detection threshold of 1 RNA copy/ml of plasma. In Figure 1, this is the parameter . If there were perfect test result conversion for all subjects (i.e. no assay ‘noise’), and further no inter-subject variability, this would reduce the smoothly varying red curve to the green dotted step function.

It is not feasible to conduct studies large enough to obtain much detail about the structure of sensitivity curves for a great number of diagnostic tests. Hence, we propose that the distribution of diagnostic delays be approximated by choosing a generic functional form which has roughly the sigmoidal structure seen in studies (our infection dating tool implements a simple cumulative normal distribution), and setting one additional parameter to capture the variability in the diagnostic delay distribution – parameter in Figure 1*.*

To calculate individual EDDIs, then, one needs to obtain estimates of the median diagnostic delays for all tests occurring in a data set, and then interpret each individual assay result as excluding some ‘non-possible’ segment of time, ultimately resulting in a final inferred interval of times during which the DDI likely exists. The prototypical situations in which one can perform dating, within this paradigm, are then when a subject:

1. tests positive at a study visit after testing negative at a previous study visit, or
2. tests positive on some component of an algorithm, and negative on another component, at a single study visit.

These calculations require that each individual has at least one negative test result and at least one positive test result. In the primitive case where there is precisely one of each, namely a negative test (with an expected diagnostic delay of ) at and a positive test with an expected diagnostic delay of at , then the interval is simply from to , with the EDDI at the midpoint. When there are multiple negative tests, at each with a diagnostic delay , and/or multiple positive tests, at each with a diagnostic delay , then each individual negative/positive test result provides a candidate Earliest Plausible Date of Detectable Infection (EP-DDI) or Latest Plausible Date of Detectable Infection (LP-DDI), respectively. The values which narrow the ‘window of detectable infection’ (i.e. the latest EP-DDI candidate, arising from the most informative negative test, and the earliest LP-DDI candidate, arising from the most informative positive test) are then the boundaries of the estimated window.

This DDI interval will not be any particular confidence interval, as the details of inter-test interval and inter-subject variability matter. When the most informative negative and positive tests are at different timepoints, the EDDI is typically the centre of a very broadly plateaued (rather than ‘peaked’) likelihood function. Given a broadly uniform prior, this can be interpreted into a round-edged plateau-like Bayesian posterior. Such a posterior, derived from an individual’s diagnostic testing history, could also serve as a prior for further analysis, if there is a quantitative biomarker available, for which there is a robustly calibrated maturation/growth curve model. We do not deal with this in the present work, but it is the subject of ongoing analysis (Pilcher *et al.*, forthcoming), and an important potential application of this framework and tool.

This simple crucial idea – interpreting the elements of a diagnostic testing history as each independently excluding some period of time from the remaining DDI interval – frees the analysis from the constraints of a pre-enumerated list of infection stages whose sequence and durations need to be assessed. It does, however, require estimation of the diagnostic delay for each assay, either by sourcing direct estimates of the diagnostic delay, or by sourcing such data for a biochemically equivalent assay. Our online HIV infection dating tool is preloaded with sensible diagnostic delay estimates for over 60 HIV assays, and users can both add new tests and provide alternative diagnostic delay estimates for those tests which are already loaded.

This interpretation of individual test results superficially appears to rely on the assumption that test results are independent (i.e. uncorrelated). However, the very factors that influence the individual sensitivity curves in Figure 1 suggest that strong correlations between results of different tests on the same person are likely. Given this, we provide below a more precise discussion of a formal inference scheme. This discussion demonstrates explicitly 1) when and how test correlation might influence the analysis (it turns out to be benign), and 2) how analysis of qualitative diagnostic testing data can interface with analysis of quantitative markers of infection stage.

# Formal Likelihood Function

It is analytically useful to specify an explicit ‘likelihood function’ – i.e. a formula for capturing the probability of seeing a data element (or set), given some hypothetical values for parameters which determine the behaviour of the underlying system, including the measurement process. This facilitates all the usual statistical manipulations for obtaining confidence intervals, Bayesian posteriors, etc. For the present application, test sensitivity curves such as those in Figure 1 are precisely the likelihood of obtaining a positive test, at a given time since infection. The likelihood of obtaining a negative result is simply 1 minus the likelihood of obtaining a positive result. As noted above, meaningful infection dating relies on having at least one negative test result and at least one positive test result.

## Classical test conversion series

To begin, we consider precisely one negative and one positive test result, arising from two subject-study interactions, at times and respectively, separated by some duration . In order to make inferences about the time of infection, we construct a likelihood function which expresses the probability of seeing these two particular results, as a function of a hypothetical infection date. This kind of likelihood (of two observations) is usually written as the product of:

* the likelihood of seeing one result (chosen arbitrarily to be considered first) given the hypothetical time of infection, and
* the likelihood of seeing the other result, given
  + the same hypothetical time of infection, and
  + the fact that the first result has already been obtained.

Defining as the actual time of infection, and and as positive and negative tests at time , respectively, this can be written as:

The details of the conditioned likelihoods, which might be complex, must necessarily be such that the two formulations are equivalent. We will focus in detail on the first formulation, as it seems more intuitively appealing when . Consider, firstly, then, the case where:

1. the correlation is very strong (qualitatively ‘positive’), as it would be if the tests performed at the two times are in fact the same test, and
2. the two test times are separated by a time that is larger than the mean diagnostic delay (and its variability).

Figure 2a shows, in solid red and green, respectively, the test sensitivity curves for the tests performed at and . Both individual-level and population-level curves (thinner and thicker lines, respectively are shown.

Figure 2a



Assumption 1 above implies that those individuals who convert rapidly on Test1 also convert rapidly on Test2. The behaviour of the full can then be understood by considering how the factor might differ from the population-averaged . The latter is what one would usually obtain from a study investigating diagnostic test performance. We now analyse the various ranges of infection time where qualitatively different situations arise:

*Values of anywhere near, or to the right of (after) :* For these ‘later’ hypothetical infection times, , we *expect* to see a negative result for the test at , and so, actually seeing the negative result provides no meaningful information on the question of whether the subject is prone to rapid or slow test conversion. Hence, no modification is implied of relative to the population average .

*Values of on the far-left end of the timeline*: For very ‘early’ hypothetical infection times, the likelihood of seeing the negative result at becomes very low. If that negative result has indeed occurred, and is not the result of a procedural error, it strongly suggests that the subject is a very slow progressor on the diagnostic marker. Even so, testing positive at (a time later than by a significant margin) is almost assured:

This analysis does not, in and of itself, handle procedural ‘test failures.’ However, in likely applications in study cohorts, unusual discrepancies would be investigated and underlying test failures would usually be detected.

*Values of within the dynamic range of the Test1 sensitivity curve*: Figure 2b considers a series of hypothetical infection times (vertical arrow labelled ) that is suggestive of the subject being a somewhat-slower-than-average progressor on the diagnostic marker. This is captured by the dashed red faster individual progression curves, indicating reduced probability. The final panel shows the full conditional likelihood.

Figure 2b











A strong correlation between tests then implies that the result for test 2 would be subject to a reduced probability of certain green individual progression curves (also dashed). However, we are specifically interested in evaluating the likelihood of seeing the particular test results at and , *given the hypothetical value of infection time, .*This samples the green curves at a time value far from their dynamic range, so that they essentially evaluate to one, just as in the preceding case of ‘early’ values of . Again, there is no meaningful difference between and the population average .

These three regimes account for the full range of values of t for which the joint likelihood is to be constructed, and means that the full joint likelihood is indeed rendered by the product of the individual unconditioned population level-likelihoods. As the curves obtain values indistinguishable from either zero or one for much of their range, this product is in turn little more than a superposition of the two curves. Figure 2c shows how this previously noted round-shouldered plateau is located relative to the test dates and population-averaged diagnostic delays.

Figure 2c

****

As indicated in **Figure 2c**, one may want to provide a pragmatic safety margin (presumably some preferred multiple of the variability parameter ) and then think of it either as either an ‘almost 100% confidence interval,’ or the region of support for a justified informative prior for subsequent analysis of a quantitative progression marker.

## Discordant results on a given study-visit

Figure 3a shows the typical ‘discordant test’ situation, where a test with a longer diagnostic delay produces a negative result and a test with a shorter diagnostic delay produces a positive result, at the same visit.

Figure 3a



Even here, though not as starkly as in the case where the two tests are conducted at significantly different times, conditioning one result on the other has relatively modest impact. Moving the hypothetical infection time to the left, the negative result becomes less likely, and the effect of the conditioning on the likelihood of seeing the second test result becomes initially more significant. It then, in turn, becomes less relevant as the considered times leave the dynamic range of Test2. The blue line is the full joint likelihood, which does not differ meaningfully from a simple product of the individual population-level likelihoods (shown in grey). The main conclusion, then, is that relative to the test date, plausible infection times are largely located between the two diagnostic delays (with some spreading due to variability).

Figure 3b shows the situation where the dynamic ranges of the tests are essentially the same. In this case, the plausible dates of infection are centred around the shared diagnostic delay of the tests, again with some spread for variability.

Figure 3b



*Outliers:* Figure 3c shows the less common situation when a more sensitive test is negative while a less sensitive test is positive. If test error has been largely ruled out, this implies an outlier situation in which the plausible test infection time likelihood is determined by the details of distributional tails and test correlation, but is nevertheless constrained to lie in the region between the two diagnostic delays. In this rare case there is less plausible scope for variability than in the typical assignment of the negative and positive results.

Figure 3c



# The Tool

The key features of our online tool for HIV infection date estimation are that:

1. Users access the tool through a website where they can register and maintain a profile which saves their work, making future calculations more efficient;=.
2. Individual test dates and positive/negative results, i.e. individual-level ‘testing histories’, not just algorithm-level diagnoses, can be uploaded in a single comma delimited text file for a group of study subjects.
3. Estimates of the relative ‘diagnostic delay’ between the assays used and the reference viral load assay must be provided, with the option of using a curated database of test properties which provides cited estimates for over 60 HIV assays.
   1. If a viral load assay’s detection threshold is known, this can be converted into a diagnostic delay estimate via the exponential growth curve model. The growth rate parameter defaults to 0.35 log10(RNA copies/ml)/day [1], but users can supply an alternative estimate.
4. Using the date arithmetic described above, when there is at least one negative test and at least one positive test for a subject, the uploaded diagnostic history results in:
   1. a point estimate for the date of first detectability of infection (the EDDI);
   2. an earliest plausible and latest plausible date of detectable infection (EP-DDI and LP-DDI); and
   3. the number of days between the EP-DDI and LP-DDI (i.e., the size of the ‘DDI interval’), which gives the user a sense of the precision of the estimate.

The logic and diagnostic test performance data required for infection dating has significant overlap with that required to calculate the residual risk of infectious material being missed by screening algorithms applied to blood products. Therefore, the online tool has a residual risk calculator built into it as well. As the question of residual risk involves additional concepts which deserve proper treatment, this aspect of the tool is discussed and presented in a separate article (Welte *et al.,* forthcoming).

## Access / User profiles

Anyone can register as a user of the tool, which can be found at <https://tools.incidence-estimation.org/idt/>. The tool saves users’ data files, and their choices about which diagnostic delay estimates to use for each assay, both of which are only accessible to the user who uploaded them. No person-identifying information is used or stored within the tool; hence, unless the subject identifiers being used to link diagnostic results can themselves be linked to people (which should be ruled out by preprocessing before upload) there is no sensitive information being stored on the system.

## Uploading diagnostic testing histories

A single data file would be expected to contain a ‘batch’ of multiple subjects’ diagnostic testing histories. Conceptually, this is a table like the fictitious example in Table 1, which records that:

* one subject (John) was seen on 10 January 2017, at which point he had a detectable vial load on an unspecified qualitative viral load assay, and a negative Western blot result (discordant tests); and
* another subject (Jill) was screened negative on a point-of-care (POC) rapid test (RT) on 13 September 2016, and then, on 4 February 2017, was confirmed positive by Western blot, having also tested positive that day on the POC RT.

Table 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Subject** | **Date** | **Test** | **Result** |
| John | 10 Jan 2017 | Qualitative VL | Positive |
| John | 10 Jan 2017 | Western blot | Negative |
| Jill | 13 September 2016 | POC RT | Negative |
| Jill | 4 February 2017 | POC RT | Positive |
| Jill | 4 February 2017 | Western blot | Positive |

In order to facilitate automated processing, the tool demands a list of column names as the first row in any input file. While extraneous columns are allowed without producing an error, there must be columns named *Subject, Date, Test* and *Result* (not case sensitive). Data in the subject column is expected to be an arbitrary string that uniquely identifies each subject.

It is fundamental to the simplicity of the algorithm that assay results be either ‘positive’ or ‘negative’. There are a small number of tests, notably Western blot and Geenius, which sometimes produce ‘indeterminate’ results (partially, but not fully, developed band structure). We now briefly reconsider Table 1 by adding the minor twist that the Western blot on subject Jill be reported as indeterminate. In this case, the data must be recorded as results on either one or both of two separate tests:

1. a ‘Test-Indeterminate’ version of the test – which notes whether a subject will be classified either as negative, or ‘at least’ as indeterminate; and
2. a ‘Test-Full’ version of the test, which determines whether a subject is fully positive or not.

There is then no longer any use for an un-suffixed version of the original test. The data from table one, with the minor change of an indeterminate result for subject Jill, is shown in Table 2, with differences highlighted. Note that even while John’s test results have not changed, his testing history now looks different, as completely negative results are reported as being negative even for the condition of being indeterminate. Jill’s indeterminate result on 4 February requires two rows to record, one to report that the test result is not fully negative (positive on ‘Western blot Indeterminate’), and one to report that the result is not fully positive (negative on ‘Western blot Full’). Once diagnostic delays are provided for these two sub-tests, the calculation of infection dates can proceed without any further data manipulation on the part of the user.

Table 2

|  |  |  |  |
| --- | --- | --- | --- |
| **Subject** | **Date** | **Test** | **Result** |
| John | 10 Jan 2017 | Qualitative VL | Positive |
| John | 10 Jan 2017 | Western blot Indet. | Negative |
| Jill | 13 September 2016 | POC RT | Negative |
| Jill | 4 February 2017 | POC RT | Positive |
| Jill | 4 February 2017 | Western blot Indet. | Positive |
| Jill | 4 February 2017 | Western blot Full | Negative |

## Provision of test diagnostic delay estimates

As described above, tests are summarised by their diagnostic delays. The database supports multiple diagnostic delay estimates for any test, acknowledging that these estimates may be provisional and/or disputed. The basic details identifying a test (i.e. name, test type) are recorded in a ‘tests’ table, and the diagnostic delay estimates are entered as records in a ‘test-properties’ table, which then naturally allows multiple estimates by allowing multiple rows which ‘link’ to a single entry in the tests table. The system’s user interface always ensures that for each user profile, there is exactly one test property estimate, chosen by the user, as ‘in use’ for infection dating calculations at any point in time.

Users need to ‘map’ the codes occurring in their data files (i.e. the strings in the ‘Test’ column of uploaded data files) to the tests and diagnostic delay estimates in the database, with the option of adding entirely new tests to the database, which will only be visible to the user who uploaded them. The tool developers welcome additional test estimates submitted for inclusion in the system-default tests/estimates.

## Execution of infection dating estimation

The command button ‘process’ becomes available when an uploaded testing history has no unmapped test codes. Pressing the button leads to values, per subject, for EP-DDI, LP-DDI, EDDI, and DDI interval, which can be previewed on-screen and downloaded as a comma delimited file. Users can specify a multiple of the diagnostic delay variability parameter () which can be added to the EP/LP-DDI interval to provide an appropriate ‘confidence interval’, or ‘range of prior support’.

## Source code, distribution, modification

The whole code base for the tool is available in a public repository on Github (<https://github.com/SACEMA/infection-dating-tool>), and so anyone can deploy their own copy of the tool, or fork the repository and make modifications, as long as the origin of the code is acknowledged and dissemination is also in open source form under the same licensing. The developers welcome contributions to the code base via ‘pull requests’. Test characteristics for the more than 60 core HIV diagnostic tests are included in the code base.

As consistent infection dating could be of interest in the study of other infections, a separate version of the system could be deployed to handle other infections, in contexts where multiple diagnostic platforms/algorithms have been used within a single data set intended for a unified analysis. This would naturally involve a completely fresh version of all data in the ‘tests’ and ‘test-properties’ tables of the tool’s database.

# Conclusion

Consistent dating of infection events across subjects has obvious utility when analysing multi-site datasets that contain different underlying screening algorithms. Consistent use of ‘diagnostic history’ information is also valuable for individual-level interpretation of infection staging at diagnosis.

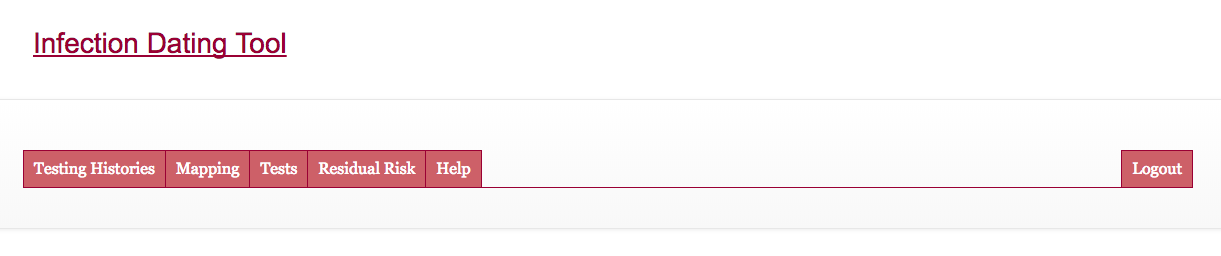
Even in intensive studies from which ‘diagnostic delay’ estimates are drawn, it is quite difficult to determine the actual date of infectious exposure. We have adopted a nomenclature based on the earliest date on which an infection would have had 50% probability of being detected using a viral load assay with a detection threshold of 1 copy per ml, and we refer to this date as the Date of Detectable Infection (DDI).

We have presented a simple logic to the interpretation of ‘diagnostic testing histories’ into ‘infection date estimates’, either as a point estimate (EDDI) or an interval (EP-DDI – LP-DDI), along with a publicly-accessible online tool that will support wider application of this logic. The DDI interval is not easily framed as a formal confidence interval or posterior percentile, due to limitations of describing test performance variability. However, this is a detail on which users of the tool can exercise their choices for safety margins.

Appendix: Web interface layout overview

Once logged in, the system presents users with four primary views/pages, accessible via links spread horizontally, tab-like, below the page header, as shown in Figure A.1. The first three are described in turn below, while the fourth (“Residual Risk”) is the subject of a separate publication (Welte *et al.*, forthcoming).

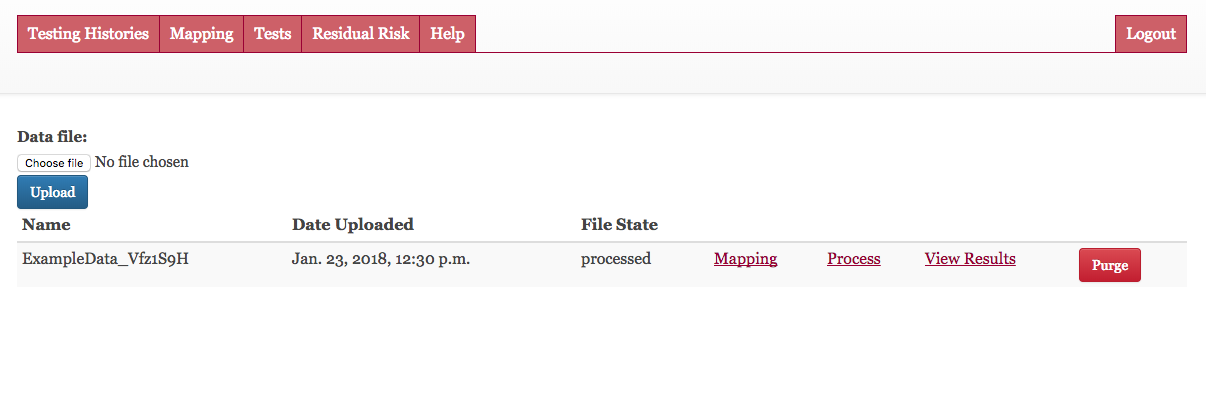
Figure A.1: Navigation



# Testing Histories

This tab (Figure A.2) allows users to locate, view and deleted previously uploaded ‘testing histories’ and to upload new ones. It is also where users trigger the action of processing the uploaded ‘testing histories’ into ‘infection dating estimates’, which can then be viewed and saved (downloaded).

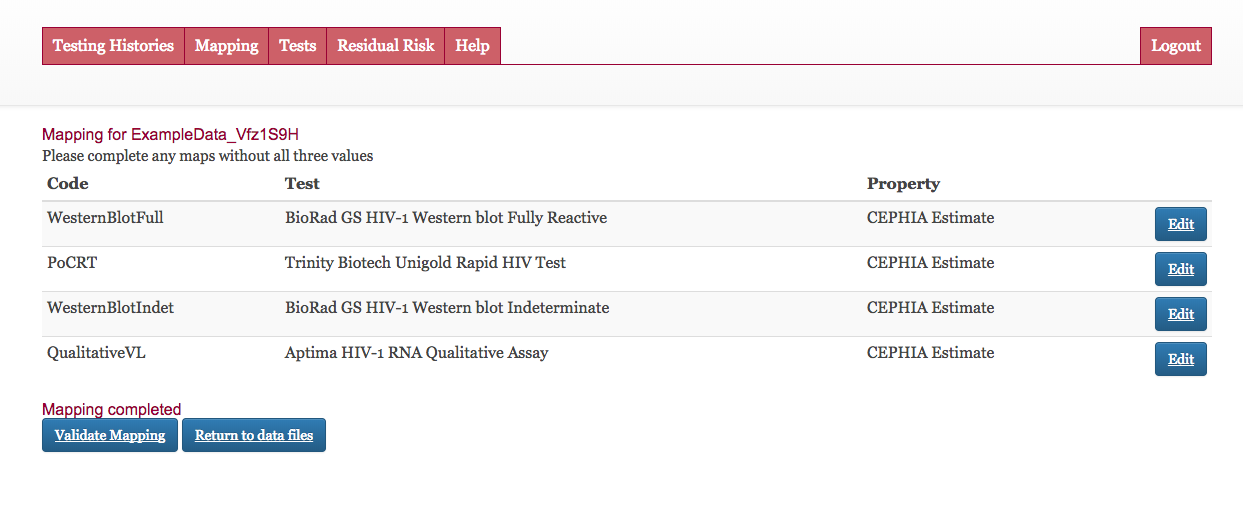
Figure A.2: Testing Histories



# Mapping

This tab (Figure A.3) allows users to link strings (alphanumeric codes) in their data files to tests in the online database, hence linking records in uploaded files to the applicable diagnostic delays.

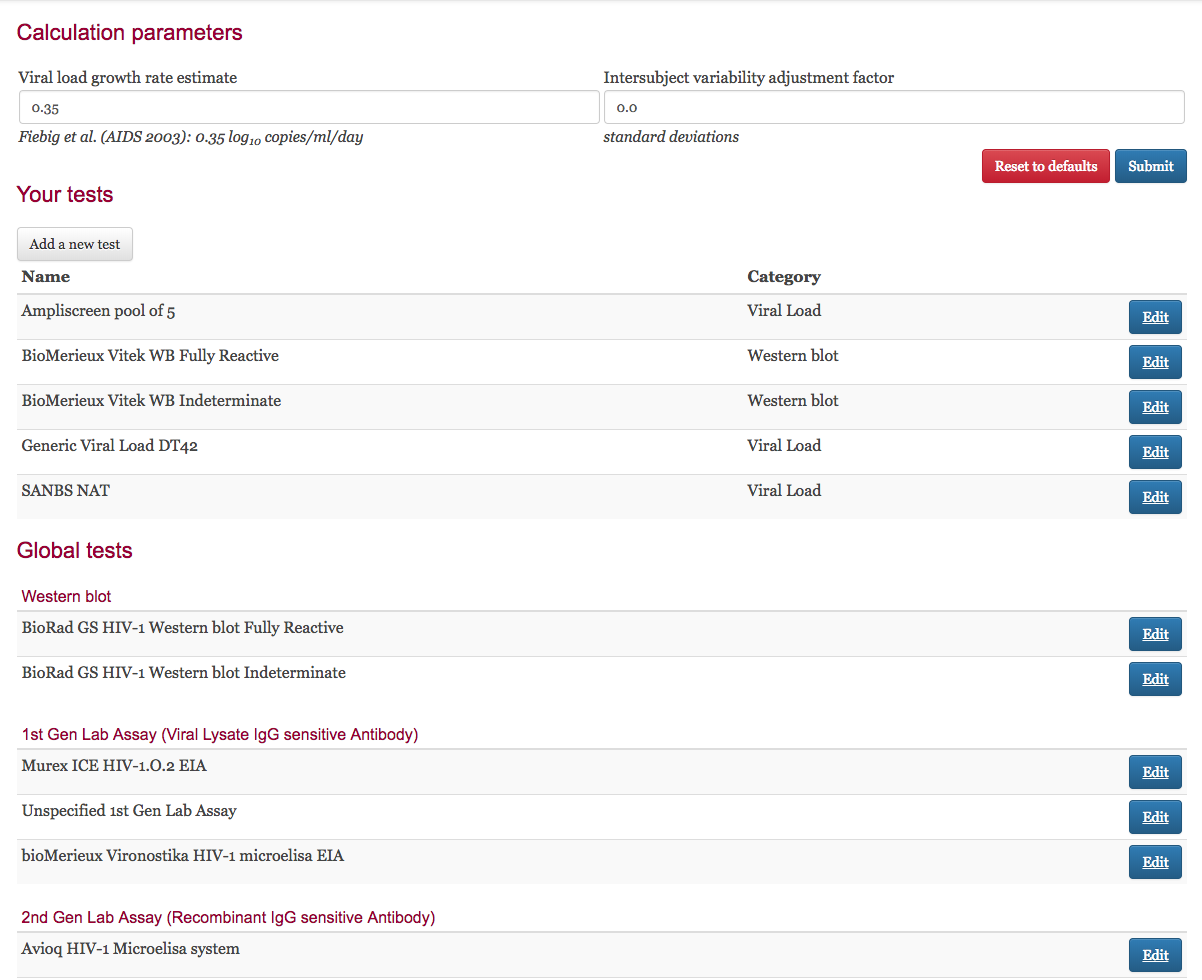
Figure A.3: Mapping



# Tests

This tab (Figure A.4) allows users to view the existing database of diagnostic tests, and to add new ones if necessary. Note that each user sees only the shared developer-maintained list of tests, plus his/her own – not those added by other users. This page further allows the user to set a viral load growth rate and to select a multiple of the intersubject variability parameter () of the relevant test’s diagnostic delay to apply to EP-DDI and LP-DDI estimates as an additional safety margin, as described in the main text.

Figure A.4: Tests



# Results

Processing can be triggered after test codes have been mapped to specific assays in the database (and test property estimates have been selected). Each file that has been uploaded on the “Testing Histories” tab has a “Mapping” link and, once mapping has been completed, “Process” link. After processing, results can be viewed and downloaded on a per-file basis (Figure A.5).

Figure A.5: Results

## 

## Database Schema

It is normal to capture study data in relational databases which record:

1. a number of concrete and abstract **entities** in **tables** (in which columns define attributes and each row records an instance of on entity, such as a person, a diagnostic assay, or a particular performance of a test together with its result), and
2. their relationships in inter-table referential constraints that link entities to each other in order to capture critical details such as whose test result is recorded, which test was used, etc.

The reader is referred to the general literature for further information on relational databases, but an intuitive grasp is sufficient for understanding this section. A few remarks on the key tables are set out below:

* **subjects**: This table captures each unique study subject, and after infection date estimation has been performed, the subject’s EDDI, EP-DDI, LP-DDI and EDDI interval size.
* **diagnostic\_test\_history**: In this table, each test performed is recorded, by linking to the subjects table, and recording a date, a ‘test code’ and a result. During the estimation procedure, a field containing an ‘adjusted date’ is populated, which records the ‘candidate’ EP-DDI (in the case of a negative result) or LP-DDI (in the case of a positive result) after the relevant diagnostic delay has been applied to the actual test date.
* **diagnostic\_tests**: This is a look-up table listing all known tests applicable to the current purposes (both system-provided and user-provided). It is conceived here for a single condition (HIV), but the existing structure can be used even if multiple agents are tested for, as long as this is clearly recorded in the metadata, and consistently used. Note that a single platform which produces multiplexed results from a single specimen must be conceived, for the present purposes, as a separate test per analyte.
* **test\_property\_estimates**: This table records diagnostic delay estimates (system and user-provided). It allows estimates per test, with system default estimates flagged.
* **test\_property\_mapping**: This table records user-specific mapping of test codes by linking (for each user of the system) each test code in the diagnostic\_test\_history table to a test in the diagnostic\_tests table as well as the specific test property estimate ‘in use’ by that user for the test in question.
* A number of subsidiary tables exist to manage users of the system and allow linking of personal data files, maps, tests, and test property estimates to specific users.

# Author contributions

AW: First draft of manuscript, overall project leadership; AW, EG and SF: conceptualisation, data curatorship, code development, writing of the manuscript; RK, MB, GM, CP: conceptualisation; AP, JG, GP, TC: code development.

# References

1. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS. 2003 Sep 5;17(13):1871–9.

2. Lee HY, Giorgi EE, Keele BF, Gaschen B, Athreya GS, Salazar-Gonzalez JF, et al. Modeling sequence evolution in acute HIV-1 infection. J Theor Biol. 2009 Nov;261(2):341–60.

3. Ananworanich J, Fletcher JLK, Pinyakorn S, van Griensven F, Vandergeeten C, Schuetz A, et al. A novel acute HIV infection staging system based on 4th generation immunoassay. Retrovirology. 2013 May 29;10(1):56.

4. CEPHIA. Consortium for the Performance and Evaluation of HIV Incidence Assays [Internet]. Available from: http://www.incidence-estimation.org/page/cephia

5. Kassanjee R, Pilcher CD, Keating SM, Facente SN, McKinney E, Price MA, et al. Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. AIDS. 2014 Oct;28(16):2439–49.

6. Murphy G, Pilcher CD, Keating SM, Kassanjee R, Facente SN, Welte A, et al. Moving towards a reliable HIV incidence test - current status, resources available, future directions and challenges ahead. Epidemiol Infect. 2017 Dec 22;145(5):925–41.

7. Kassanjee R, Pilcher CD, Busch MP, Murphy G, Facente SN, Keating SM, et al. Viral load criteria and threshold optimization to improve HIV incidence assay characteristics. AIDS. 2016 Sep;30(15):2361–71.

8. Grebe E, Welte A, Hall J, Keating SM, Facente SN, Marson K, et al. Infection Staging and Incidence Surveillance Applications of High Dynamic Range Diagnostic Immuno-Assay Platforms. JAIDS J Acquir Immune Defic Syndr. 2017 Dec;76(5):547–55.

9. Kassanjee R, McWalter TA, Bärnighausen T, Welte A. A New General Biomarker-based Incidence Estimator. Epidemiology. 2012 Sep;23(5):721–8.

10. Owen SM, Yang C, Spira T, Ou CY, Pau CP, Parekh BS, et al. Alternative algorithms for human immunodeficiency virus infection diagnosis using tests that are licensed in the United States. J Clin Microbiol. 2008;46(5):1588–95.

11. Masciotra S, McDougal JS, Feldman J, Sprinkle P, Wesolowski L, Owen SM. Evaluation of an alternative HIV diagnostic algorithm using specimens from seroconversion panels and persons with established HIV infections. J Clin Virol. 2011 Dec;52 Suppl 1(SUPPL. 1):S17-22.

12. Delaney KP, Hanson DL, Masciotra S, Ethridge SF, Wesolowski L, Owen SM. Time Until Emergence of HIV Test Reactivity Following Infection With HIV-1: Implications for Interpreting Test Results and Retesting After Exposure. Clin Infect Dis. 2017 Jan 1;64(1):53–9.

13. Konrad BP, Taylor D, Conway JM, Ogilvie GS, Coombs D. On the duration of the period between exposure to HIV and detectable infection. Epidemics. 2017 Sep;20:73–83.

14. Kassanjee R. Characterisation and Application of Tests for Recent Infection for HIV Incidence Surveillance. Vol. PhD, School of Computational and Applied Mathematics. [Johannesburg]: University of the Witwatersrand; 2014.