A Method to Derive the Time of Onset of Infection from Serological Findings[1]

Friedrich Steimann[1] and Michael Hayde[2]
[1]Institute of Medical Informatics, University of Hildesheim, Germany
[2]University Children's Hospital, University of Vienna, Austria

Abstract: Among the diagnostic problems that require a retrospective assessment of the time an event occurred is that of screening for primary infection with *Toxoplasma gondii* during pregnancy. We suggest a method to derive the possible times of onset of infection from a small sequence of serological samples by matching them against the knowledge about possible courses of infection. Special care is taken to properly address the relative change (gradient) of consecutive samples, a nontrivial problem when reasoning about sparsely sampled courses. To demonstrate the practicability of our approach we conducted an evaluation based on the samples of 394 pregnancies selected at random from our toxoplasmosis database; we could show an accuracy of 95.7%.

Keywords: serodiagnosis, temporal reasoning, reasoning about relative change, onset of infection, congenital toxoplasmosis

1. Introduction

Certain diagnostic problems require an assessment of the time some triggering event occurred. This is problematic if the event itself remains obscure (asymptomatic) or is not an issue at its time of occurrence. Such is usually the case when a woman becomes infected with the parasite *Toxoplasma gondii*. If that woman is also pregnant, she risks transmitting the pathogen to her unborn child, seriously endangering its life and health. In some countries, screening programs have therefore been implemented to routinely test pregnant women for postconceptional infection with *Toxoplasma gondii*. Because this procedure has to differentiate postconceptional from preconceptional infections, it inevitably involves some kind of temporal reasoning.

The method we present performs the required reasoning by deriving all possible times of onset of infection from the available serological evidence. It does so by matching the findings against the serological knowledge about possible courses of infection. We take special care to properly address the problem of reasoning about the relative change between consecutive findings (which is inherently difficult with sparse sampling) and show the theoretical soundness of our solution. That it is also practically efficacious is demonstrated in an evaluation based on the samples of 394 pregnant women selected at random from our routine screening database. To investigate the dynamic performance of our method we simulate its behaviour in prospective employment.

2. Problem

An estimated more than 20% [Lappalainen 93] of the world population and countless animals are infected with the parasite *Toxoplasma gondii*. Its main mode of transmission is by ingestion of oocysts, toxoplasma cells excreted by cats and spread around by flies and other insects, and by consumption of raw meat from infected animals [McCabe 83].

Once acquired, the pathogen spreads throughout the body via lymphatics and the bloodstream and infests tissue cells, mostly those of the central nervous system, heart, and muscles. After this *acute* phase of infection (also referred to as *recent infection*), cell-mediated and humoral immune response eliminate the parasite from the bloodstream, leaving viable organisms only to

[1] A shortened version of this paper appeared in *Methods of Information in Medicine* 36:1 (1997) 51–58

persist encysted in the host tissue. This latter *chronic* or *latent* stage of infection is assumed to last for the whole lifetime of the host. [McCabe 83, Desmonts 85]

Toxoplasmosis is the clinical disease associated with an infection with Toxoplasma gondii. It is not an obligatory consequence of acute infection—only in about 10% of the population postnatal infection is symptomatic. Typical signs of *postnatal toxoplasmosis* are those of mononucleosis, namely malaise, fever, headache and swollen lymph nodes. Because it is generally well controlled by the immune system, toxoplasmosis is not life-threatening for the immunocompetent person. More recently, however, postnatal toxoplasmosis has gained in significance, as the number of immunocompromised (such as transplant recipients and other immunosuppressed patients) and immunodeficient (AIDS!) individuals increases.

Prenatal toxoplasmosis, on the other hand, has longer been a matter of concern: transplacentally transmitted pathogens causing *congenital toxoplasmosis* in the foetus can destroy the unborn's developing tissue and cause irreparable damage. This fact gives rise to toxoplasma screening of pregnant women; more on this below.

2.1. Serological tests, courses of infection and serodiagnosis

For practical reasons, diagnosis of toxoplasma infection is largely based on indirect serological tests detecting specific toxoplasma antibodies. These tests vary in the type of antibody and in the quality they respond to. Most frequently used are the tests for the determination of IgG and IgM antibody concentrations. More recently, tests to measure IgG antigen-binding avidity [Lappalainen 93] and IgA [Bessièrs 92] have come to support serodiagnosis.

Our laboratory, the Toxoplasmosis Laboratory of the University Children's Hospital in Vienna, Austria, currently uses three tests:

1. Sabin-Feldman dye test (DT), the World Health Organization reference test measuring mostly specific IgG antibodies;
2. immunosorbent agglutination assay (ISAGA, Bio-Merieux) for the detection of IgM antibodies; and
3. solid-phase enzyme immunoassay for the determination of IgG antibody avidity.

All test results are obtained and interpreted in their quantitative form, i.e., the DT as a titer in steps of fourfold dilution ranging from 0 to 1 : 65536, IgM ISAGA as an index in the range of 0–12, and IgG avidity as a percentage.

In response to primary contact with antigens, first specific IgM and then IgG antibodies are produced by the immune system. Repeated testing makes visible the course of infection: Figure 1 is an adaptation of courses published in [Desmonts 85, Bessières 92, Lappalainen 93]. Different tests of the same antibody quality display different courses; serodiagnosis therefore heavily relies on knowing the properties of the employed tests.

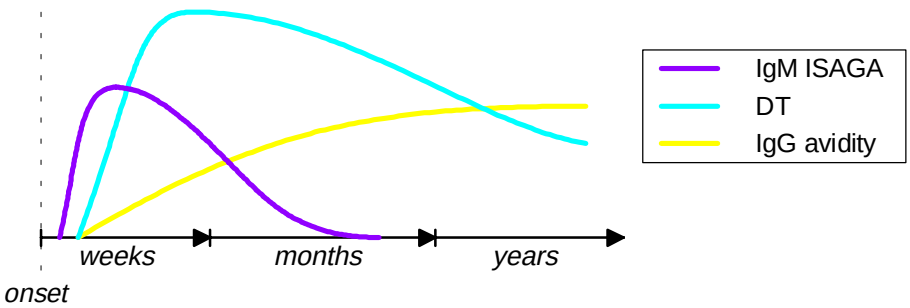

Figure 1: Typical course of antibody production after primary infection with Toxoplasma gondii as observable through various tests (adapted from [Desmonts 85, Bessières 92, Lappalainen 93])

The pre-eminent property of the DT is its high sensitivity and specificity: positive titers prove toxoplasma infection while a negative (zero) titer excludes it. Unfortunately, its capability of discriminating acute from chronic infections based on a single serum is limited: individual immune response to toxoplasma antigen varies so considerably in both strength and speed that reliable assessment of acuteness can only—if at all—be based on the relative change of paired sera. Originally intended as a confirmatory test, the IgM ISAGA has only potential to rule out acute infection: whereas high indices can occur with both acute and latent infections (high titer persistencies), a negative IgM is never found with an acute infection. The IgG avidity test, a new technique for the measurement of the antigen-binding avidity (functional affinity) of IgG, distinguishes low-affinity antibodies at an early stage of infection from those with a higher binding affinity reflecting pre-existing immunity [Lappalainen 93]. Other than the DT and the IgM ISAGA, its outcome is monotonically increasing with the age of infection, i.e., in an individual patient each finding is reversibly related to one and only one time after infection.

Typical courses of reinfection with Toxoplasma gondii have not been published—apparently, repeated contact with the parasite leaves the antibody concentration unaltered. This is maybe because cysts leak and so emit antigens to the bloodstream, thus keeping immune activity on a certain level. Be it as may, a significant increase in antibody concentration appears to be symptomatic of acute infection and acute infection only. This fact is heavily relied on in serodiagnosis.

2.2. Congenital toxoplasmosis

Because the foetus's immune system is still undeveloped, transplacental transmission of the parasite may result in congenital toxoplasmosis. Congenital toxoplasmosis is known to cause severe damage ranging from foetal death and stillbirth through hydrocephalus to clinically healthy newborns with an 80% chance of developing ocular toxoplasmosis and blindness in adulthood. It is therefore to be avoided.

Practically, only mothers with primary infection acquired during pregnancy (so-called *postconceptional* or *gestational* primoinfections) are at risk[2]. If infection is acquired before conception (so-called *preconceptional infection*), no practical risk of transmission exists [McCabe 83]. Figure 2 visualizes the terminology.

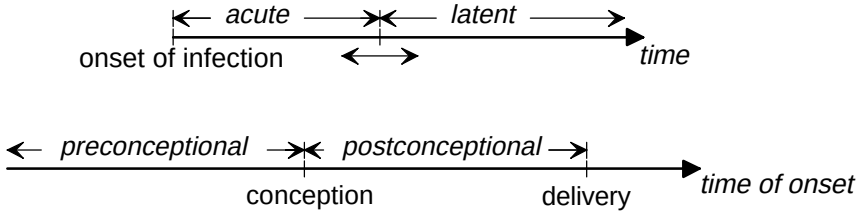

Figure 2: Naming of infections

In the context of congenital toxoplasmosis it is thus not the acuteness of infection that matters, but its time of onset relative to the date of conception. It is important that acute and postconceptional infection are not confused, nor latent and preconceptional infection. For example, a woman with evident toxoplasmic lymphadenopathy (diseased lymph nodes diagnosed by, e.g., biopsy, as due to acute toxoplasma infection) who becomes pregnant is acutely, yet preconceptionally, infected and therefore bears no risk of transplacental transmission.

[2] which further depends on the gestational age [Remington 90]

2.3. Screening of pregnant women

Routine screening for toxoplasma infection as implemented in France and Austria aims at preventing congenital toxoplasmosis [Aspöck 92]. To achieve prevention, all pregnant women with postconceptionally acquired infection must be identified and treated.

The efficacy of toxoplasma screening has been demonstrated in several studies, for example [Aspöck 92, Lappalainen 93]. With early diagnosis and treatment, the incidence of congenital toxoplasmosis can be reduced from 2.4–7‰[3] to less than 0.1‰. These figures imply that only few cases of postconceptional infection are overlooked, which means that the diagnostic procedure currently employed in screening is highly sensitive. Figures from our toxoplasmosis laboratory, however, suggest that it is not very specific, as of approximately 180 treatments per 10,000 births 110–156 cases (61%–87%) are treated in excess of what would have been necessary.

This observation is corroborated by theoretical considerations. As noted in [Aedes 91], current screening tests of recent infection are unlikely to be both sensitive and predictive. The reason lies in the wide spread of individual courses of infection, plaguing the onset determination with inevitable inaccuracy. This inaccuracy is also reflected in the meaning of the diagnostic terms *recent* and *latent infection*: if, for example, recent infection means acquired during the last six months, this diagnosis is unsatisfactory for the first 24 weeks of gestation, because postconceptional infection can neither be excluded nor confirmed.

3. Method

Rather than striving to reach impossible precision, methodical information processing should recognize the inaccuracy as an inherent part of the problem and its solution. Objective serodiagnosis then has to differentiate at least three possible situations: *preconceptional infection*, *postconceptional infection*, and *uncertain*. As will be shown, such can be done with high accuracy. The remaining downside, that the diagnosis *uncertain* is not satisfactorily specific, may appear unfortunate, yet has to be put up with unless better serological tests become available.

The method is based on the following considerations. Biological inter-patient variability hinders us from accurately determining the time of onset on the basis of sparse[4] serological samples. Yet, this variability is within limits, and our knowledge about these limits allows us to exclude certain times of onset, and this with certainty. For example, as we know that the DT reaches its peak titers—which are always greater than 1:256—eight to nine weeks after onset of infection, a titer as low as 1:64—without saying much about the true onset of infection—immediately rules out an infection that is eight weeks old. By combination of such indirect evidence obtained from different tests and sera we can encircle the possible times of onset on which the final diagnosis is based.

3.1. Reasoning about the absolute course

The exclusion of possible times of onset can be formalized as follows. Let x be a serological test. The results of x depend on the patient and the time of sampling. x may thus be viewed as a (pointwisely defined, partial) function mapping the set of patients P and time T (the real line) into x's range, V_x, the set of possible test results;

$$x : P \times T \to V_x .$$

For some patient $p_i \in P$ and time $t_0 \in T$, $x(p_i, t_0)$ then denotes the sample (test result) taken of p_i at t_0. Further let

[3] wide spread partly due to different decades and regions from which figures stem

[4] Casually speaking, sampling is sparse if the original signal cannot be reconstructed from the samples.

$$c_x^- : T \to V_x \text{ and } c_x^+ : T \to V_x$$

be two functions such that $c_x^-(t_0)$ is the least possible outcome of test x at t_0 after the onset of infection and $c_x^+(t_0)$ is the greatest. (Without loss of generality we assume that $c_x^-(t_0) = c_x^+(t_0) = 0$ for all $t_0 \leq 0$, which is in accordance with serology.) For any given sample $x(p_i, t_0)$, t_Ω is then a *possible time of onset of infection* only if

$$x(p_i, t_0) \in [c_x^-(t_0 - t_\Omega), c_x^+(t_0 - t_\Omega)].$$

In particular, if $x(p_i, t_0) < c_x^-(t_0 - t_\Omega)$ or $x(p_i, t_0) > c_x^+(t_0 - t_\Omega)$, t_Ω is definitely ruled out as a time of onset. By testing this condition for all $t_\Omega \in T$, intervals of possible times of onset can be identified, as shown in Figure 3. The corresponding diagnosis is obtained by relating the derived intervals with the date of conception; it can be looked up in Table 1.

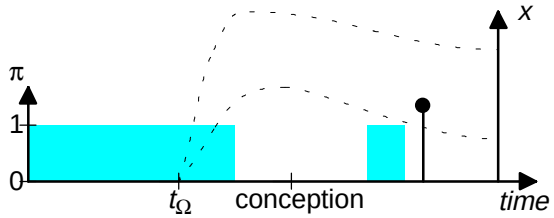

Figure 3: Deriving possible times of onset; dot and scale on the right mark the finding, grey boxes and scale on the left indicate the possible times of onset (π = 1); dashed lines represent $c_x^-(t - t_\Omega)$ and $c_x^+(t - t_\Omega)$; in this particular case, times of onset both before and after conception are possible, the diagnosis thus being *uncertain* (see Table 1)

Table 1: Mapping possible times of onset to diagnoses

	ONSET POSSIBLE		DIAGNOSIS
	BEFORE CONCEPTION	AFTER CONCEPTION	
AT LEAST ONE POSITIVE FINDING	*yes*	*no*	*preconceptional infection*
	no	*yes*	*postconceptional infection*
	yes	*yes*	*uncertain*
	no	*no*	*inconsistent data*
ALL FINDINGS NEGATIVE			*no infection*

Table 1 reflects the clinical perspective of serodiagnosis. Theoretically, if all DT samples remain negative during pregnancy, infection is possible only postpartum, which is also after conception. Within the context of screening, however, possible infections after delivery are of no concern (unless they coincide with another pregnancy), so that the diagnosis of a DT seronegative patient is *no infection* (before or during this pregnancy). Hence the subdividing of Table 1.

As suggested by the example of Figure 3, serodiagnosis based on a single sample is rather unspecific. To reduce the degree of indecision, clinical practice resorts to the combination of evidence obtained by the employment of additional tests and follow-up serology.

But how is the evidence to be combined? Because a single sample cannot prove a hypothesized time of onset, but only exclude it, evidence obtained from different samples must be combined in a conjunctive fashion. More specifically, if one sample excludes infection at a certain time, it overrules other samples allowing an infection at that time. The possibility π of time t_Ω being the onset of infection regarding all samples is thus obtained by joining the possibilities derived from the single samples with a logical *and*. Formally, for *n* samples of patient p_i

$$\pi(t_\Omega) = \bigwedge_{1 \le j \le n} c^-_{x_j}(t_j - t_\Omega) \le x_j(p_i, t_j) \le c^+_{x_j}(t_j - t_\Omega) \,.^{5}$$

The following examples schematically demonstrate the derivation of diagnoses from typical findings.

Example: (postconceptional seroconversion)

Initially, when no information is available, the onset of infection seems equally possible at all past and future times. The corresponding distribution of possible times of onset is shown in Figure 4 a). With the first DT sample, however, the picture changes drastically. If it is negative, the patient is not yet infected so that only future infection remains possible, a fact that is reflected in the distribution of Figure 4 b). (If it were positive, the patient would already be infected, and future primoinfection would be excluded.) A subsequent positive finding excludes later onset of infection, so that possible onsets are restricted to times between the samples, as shown in Figure 4 c).

◊

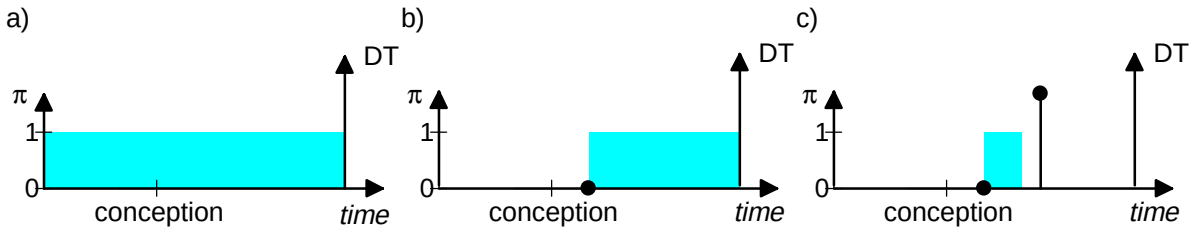

Figure 4: Distribution of possibility of times of onset a) when nothing is known, b) after a negative sample, and c) after a seroconversion

Example: (preconceptional infection)

If the first DT performed during pregnancy is positive, infection is evident, yet may have equally possibly been acquired at times before and after conception, as shown in Figure 5 a). Note that the titer height alone is not indicative of the recency of infection—both high and low titers can be found in acute and latent infections. An IgM ISAGA performed on the same serum may provide a different picture: as Figure 5 b) demonstrates, if the sample is negative, the patient is either not (yet) infected or infection is latent. Only in combination with the DT of Figure 5 a) can the infection be dated back to times before conception, as shown in Figure 5 c).

◊

[5] Note that neither the tests x_j nor the sampling times t_j need be pairwisely different, as the same test may be (and usually is) repeated at later times and different tests may be performed at the same time.

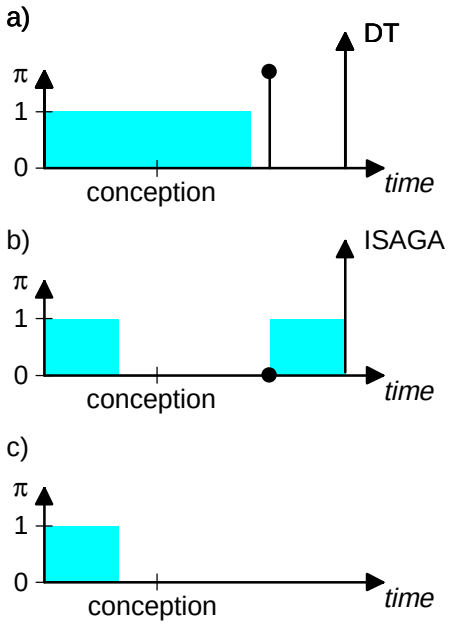

Figure 5: Same as Figure 4 for
a) a positive DT,
b) a negative IgM ISAGA, and
c) the combination of both

3.2. Reasoning about relative change

Conjunctive combination of evidence is logically consequential—yet, alone it is insufficient to challange human expert interpretation of paired sera as it fails to exploit the information immanent in the relative change of successive findings of the same test. Indeed, if the spread of $c_x^-(t)$ and $c_x^+(t)$ is only wide enough, sequences of findings suggestive of a falling development match with the rising phase of the course (and vice versa), a problem that is visualized in Figure 6 a). This theoretical consideration was confirmed by a first retrospective evaluation of the approach which showed that this problem caused 86% of the total misclassifications [Steimann 94]. Note that constraining the divided difference of two consecutive samples is no remedy: if sampling intervals are long, nothing can be said about the gradient of the course at either sampling time, so that reasoning about the gradient is invariably speculative. An example of two significantly different courses competing in the explanation of a pair of samples is given in Figure 6 b).

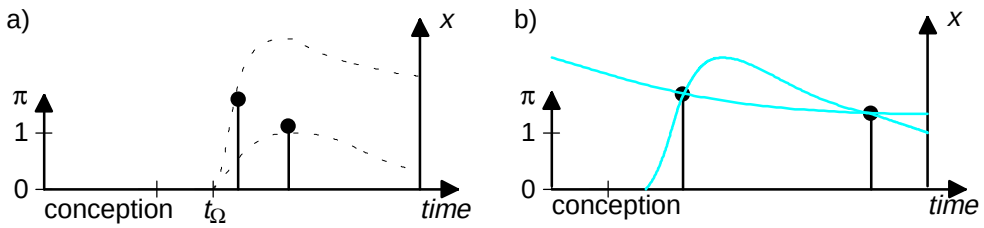

Figure 6: Diagnosis based on paired sera; a) possible misinterpretation of findings suggestive of a latent (and preconceptional) infection due to the wide spread of possible courses; b) long intervals between samples disallowing assumptions about the gradient (first derivative, momentary rate of change) of findings

These problems led to the employment of so-called *explaining courses* [Steimann 95a, Steimann 95c], hypothetical continuous-time functions meeting all the findings of a given patient and test so that their sampling at the given times would produce the findings, hence explaining them. Because the explaining courses are continuous, they can be differentiated, and the result

can be matched with a pair of functions constraining the first derivative of the course of infection.

Let $d_x^-(t)$ and $d_x^+(t)$ be two such functions constraining the first derivative of possible courses of infection, i.e., $d_x^-(t_0)$ is the least gradient possible at t_0 after onset and $d_x^+(t_0)$ is the greatest. Again we assume that $d_x^-(t_0) = d_x^+(t_0) = 0$ for all $t_0 \leq 0$. Given the sequence of samples $\langle x(p_i, t_1), ..., x(p_i, t_n)\rangle$, t_Ω is a possible time of onset only if there exists a differentiable function $x_e : T \to V_x$ such that

$$x_e(t_j) = x(p_i, t_j) \quad \text{for all } 0 \leq j \leq n \tag{1}$$

and

$$d_x^-(t - t_\Omega) \leq x_e'(t) \leq d_x^+(t - t_\Omega) \quad \text{for all } t, \tag{2}$$

where $t_j = t_\Omega$ for $j = 0$ and $x(p_i, t_\Omega) = 0$ have been defined to reflect the fact that a test must be zero (negative) at the time of onset.

Note how this approach elegantly solves both problems outlined in Figure 6. In Figure 6 a) the derivative of every explaining course must be negative at some time between the first and the second sample. This is in contrast to the fact that during the initial (acute) phase of infection, both $d_x^-(t)$ and $d_x^+(t)$ will be non-negative, thus rendering recent times of onset impossible. On the other hand, after the acute phase of infection both $d_x^-(t)$ and $d_x^+(t)$ will be non-positive, thus allowing an earlier onset of infection in the case of Figure 6 a) and both a latent and a recent infection in Figure 6 b), the latter with a peak between samples.

From the computational standpoint it is important to know that a suitable explaining course —if it exists—can be derived systematically from the samples and $d_x^-(t)$ and $d_x^+(t)$, thereby allowing a simple test for the possibility of any hypothesized time of onset. In fact, an explaining course with the time of onset t_Ω satisfying (1) and (2) exists if and only if

$$\int_{t_j - t_\Omega}^{t_{j+1} - t_\Omega} d_x^-(t)dt \leq x(p_i, t_{j+1}) - x(p_i, t_j) \leq \int_{t_j - t_\Omega}^{t_{j+1} - t_\Omega} d_x^+(t)dt \quad \text{for all } 0 \leq j < n. \tag{3}$$

The proof can be found in the Appendix.

4. Acquisition of the serological knowledge

One problem that remains to be solved is the acquisition of the serological knowledge in terms of the functions bounding naturally occurring courses of infection and their derivatives. Standard deviations from the published courses of Figure 1 are not appropriate, because for the presented method to be 100% sensitive it requires an assessment of all possible courses. We therefore decided to found this assessment on serological data from the evident acute primoinfections documented at our laboratory.

Briefly, $c_x^-(t)$ and $c_x^+(t)$ are acquired by reversing the above onset determination process. For this purpose, a clinician was asked to assess an interval $[t_{p_i}^-, t_{p_i}^+]$ of possible times of onset of infection for each of $n = 42$ acutely infected cases p_i, based on the findings $x(p_i, t_j)$ obtained through the DT, IgM ISAGA and IgG avidity, as shown in Figure 7 a). From this assessment follows (by aligning the specified times of onset to $t = 0$) that for each test x, $x(p_i, t_j)$ is a possible finding in the interval $[t_j - t_{p_i}^+, t_j - t_{p_i}^-]$ after the onset of infection, as depicted in Figure 7 b).

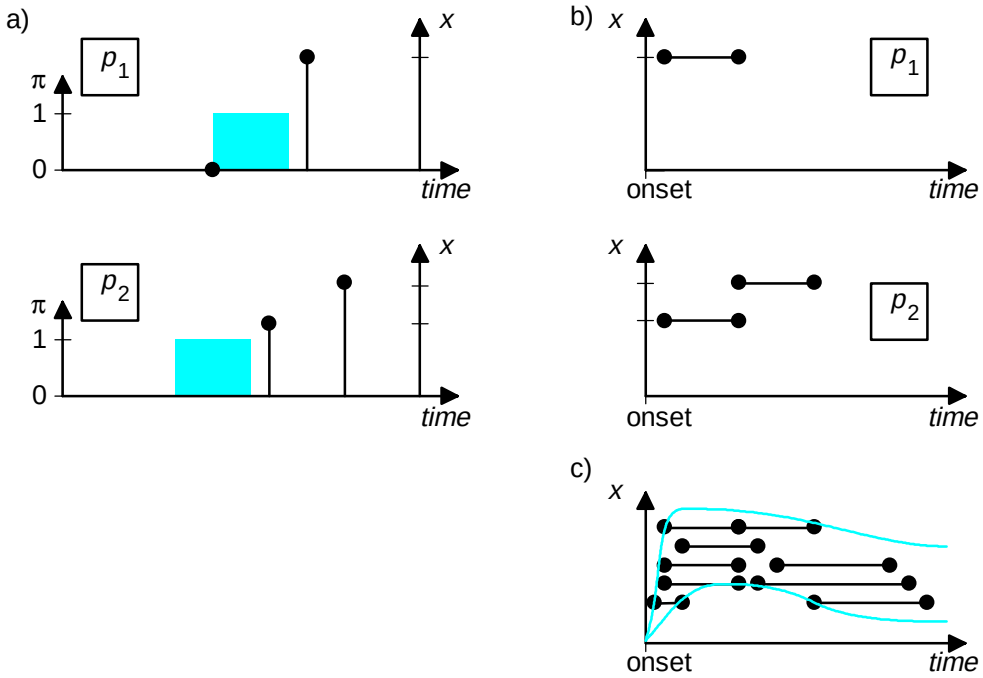

Figure 7: Acquisition of possible courses of infection
a) assessments of possible times of onset for evident acute infections by an expert
b) translation of these assessments to intervals of possible findings
c) superposition and envelope of possible findings

The bounding functions $c_x^-(t)$ and $c_x^+(t)$ are then defined for each test x so that they include all translated findings of that test, as shown in Figure 7 c).

To derive the bounding derivatives we proceeded analogously, only that we based them on the divided difference of each pair of consecutive samples, assigned to the center of the sampling interval.

Applied to the samples of DT and IgM ISAGA this procedure led to the courses shown in Figure 8. It should be clear that the courses have a subjective component, as they reflect the clinician's conception of possible courses on which he bases his routine diagnosis, explicated in his assignment of intervals. Because the true time of onset of infection is only known in very few cases, it is possible that these courses do not comply with reality.[6] The bounding functions of IgG avidity reflect a modified[7] form of the diagnostic criteria published in [Lappalainen 92] as presented in Table 2. Note the coarseness of the rules revealed in their graphical translation; it will be returned to in the discussion.

[6] The fact that neither of the courses takes the delayed onset of measurable immune response into account hints at such a discrepancy.

[7] Modifications are minor and partly reflect deviating criteria published by the same authors.

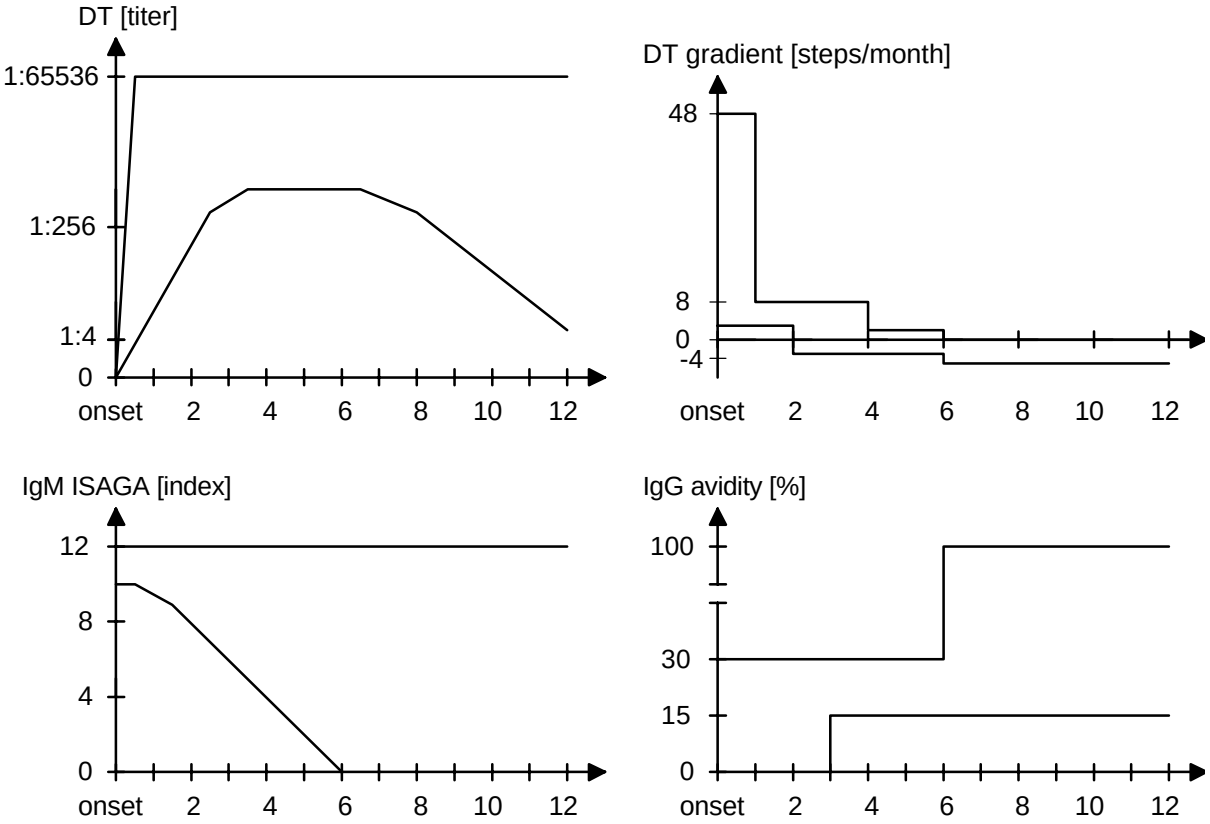

Figure 8: Bounding functions of possible courses of acute primary infection as observable through DT, IgM ISAGA and IgG avidity (the latter adapted from decision criteria published in [Lappalainen 92]); the slope of the DT's explaining courses is additionally restricted by minimum and maximum titer gradients; time in months

Table 2: Interpretation of IgG avidity results

IGG AVIDITY	CLASSIFICATION	INTERPRETATION
≤ 15%	*low*	infection acquired within last three months
16–29%	*borderline*	no statement possible
≥ 30%	*high*	infection more than six months ago

5. Evaluation

Based on the graphs of Figure 8 and the diagnostic criteria of Table 1 we implemented a computer program called ONSET that derives the diagnoses automatically from serological data recorded in a database. To evaluate ONSET's diagnostic performance we selected the samples of 1,000 women having follow-up serology, chosen at random from our screening database. Out of these, 606 cases remained seronegative throughout pregnancy; because their diagnosis *no infection* (*cf.* Table 1) is trivial, they are not included in this evaluation. The remaining presented with the number of samples shown in Table 3.

Table 3: Number of samples available from the 394 cases involved in the evaluation

TEST	1ST SERUM	2ND SERUM	3RD SERUM	4TH SERUM	5TH SERUM	TOTAL
DT	394	394	70	19	3	880
IGM ISAGA	208	164	36	11	2	421
IGG AVIDITY	102	59	8	2	0	171

5.1. Retrospective evaluation

All 394 cases were diagnosed by ONSET and by a clinician. The clinician was urged not to question the correctness of individual findings (e.g., to blame aberrant findings on measurement error and thus ignore them) and declare all cases of which he felt that data was erroneous as *inconsistent data*. The only exception that both the clinician and ONSET were allowed to make was to regard an increase in DT by one step (which is a frequent consequence of test inherent imprecision colliding with the slope restriction of the latent stage of infection, see Figure 8) as constant.

Table 4: Performance of ONSET compared to a clinician

ONSET	CLINICIAN				
	POSTCONCEPTIONAL	UNCERTAIN	PRECONCEPTIONAL	INCONSISTENT DATA	TOTAL
POSTCONCEPTIONAL	3	1	–	–	4
UNCERTAIN	1	30	3	–	34
PRECONCEPTIONAL	–	7	335	2	344
INCONSISTENT DATA	1	–	3	8	12
TOTAL	5	38	341	10	394

The overall accuracy of ONSET on the selected cases is 95.7%. Even more encouraging is the fact that only 5 of the 18 deviating diagnoses were judged unacceptable by the clinician. Of those,

- all three cases classified *uncertain* by ONSET and *preconceptional* by the clinician had low DT titers (1 : 1024 or 1 : 256) at the end of pregnancy which have also been found in evident acute infections and were thus included in the bounding functions;
- one postconceptional infection was considered inconsistent by ONSET because a rapid increase (from 1 : 64 to 1 : 65536 within twelve days, corresponding to a slope of more than twelve steps per month) was not followed by higher titers, an observation that is not compatible with the slope restriction of Figure 8 which would require a further titer rise (for practical reasons titration usually ends at 1 : 65536 so that higher concentrations, even if present, are not observed); and
- one preconceptional infection was classified as inconsistent by ONSET because the DT decayed slightly faster than tolerated by the slope restriction.

Contrasted with the diagnostic performance of ONSET published in [Steimann 94], its current implementation shows significantly increased congruence with the clinician's diagnoses. Most notably, no false classifications are due to the disregard of relative change, which presented a major deficiency of the earlier version. The result is the more impressive as the courses were updated to represent a wider range of acute infections, reflecting more general (and hence less specific) serological knowledge. In numbers, employment of the slope restriction via explaining courses influences ONSET's diagnoses as shown in Table 5.

Table 5: Impact of slope restriction on ONSET's diagnoses

DIAGNOSIS	WITHOUT SLOPE RESTRICTION		WITH SLOPE RESTRICTION
postconceptional	3	1 →	4
uncertain	80		34
preconceptional	303	44 →	344
inconsistent data	8		12

total	394	394

5.2. Simulated prospective evaluation

The retrospective nature of this evaluation allows both the clinician and ONSET to fully exploit the availability of follow-up data. In everyday practice, however, a diagnosis must be made after each serum, and this diagnosis is subject to confirmation or change after each follow-up. To investigate the prospective performance of ONSET we simulated its routine employment by diagnosing the same 394 cases incrementally, i.e., serum after serum.

To trace the dynamics of ONSET's diagnoses, we determined the number of changes from one diagnostic class to another after each follow-up serum. The results presented in Table 6 show that a high fraction (55.5%) of all preconceptional infections is classified as *uncertain* after the first serum. Differentiation is highest with the second serum: almost half of all diagnoses made at this stage are refinements (changes from *uncertain* to *preconceptional* or *postconceptional infection*) of their previous diagnoses.

Table 6: ONSET's diagnoses after each serum. The first term of each sum counts the number of cases whose diagnoses remain unchanged on follow-up, while the second counts the ones that change. Numbers on arrows state the number of refinements to *pre-* or *postconceptional* with each new serum.

DIAGNOSIS	AFTER 1ST SERUM	AFTER 2ND SERUM	AFTER 3RD SERUM	AFTER 4TH SERUM	AFTER 5TH SERUM
postconcept.	2+0	4+3	2+0	1+0	
uncertain	41+195	36+5	10+2	5+0	1+0
preconcept.	152+2	338+0	48+0	11+0	2+0
inconsistent	2+0	8+0	8+0	2+0	
total	197+197	386+8	68+2	19+0	3+0

5 (*postconcept.* → *uncertain*); 186 (*uncertain* → *preconcept.*); 4 (*uncertain* → *preconcept.*); 2 (*uncertain* → *preconcept.*)

Disappointingly, Table 6 does not provide any evidence that ONSET could save therapies. Quite to the contrary, if all initially uncertain cases were treated, the treatment rate of the chosen cohort would be more than ten times higher than the average rate at our laboratory. On the other hand, Table 6 suggests that 45.3% (220) of all follow-up examinations of seropositive mothers could be saved if ONSET's diagnoses were trusted. In particular, the prospective evaluation shows that while a second serum makes a difference in 50% of all cases, the third and all further sera are only seldom useful.

The high degree of indecision after the first serum is mostly due to the fact that for the slope restriction to catch at least two samples ("paired sera") are required. Some of the initially uncertain cases would have been classified *preconceptional* if a (mostly negative) IgM had been available. However, considering the fact that positive IgM results allow no absolute statement about the onset of infection, the results impressively justify the serological necessity of follow-up serology after a first positive sample.

It remains to be noted that the fraction of *uncertain* cases increases with the number of sera taken, reflecting the laboratory's search for evidence that would differentiate these cases.

6. Discussion

Despite its demonstrated diagnostic competence, the ONSET approach leaves considerable room for improvements. Firstly, the imprecision of its input should be taken into account. As touched on earlier, due to test-inherent sources of inaccuracy a DT titer of $1 : 4^n$ may indeed be compatible with one of $1 : 4^{n-1}$ or one of $1 : 4^{n+1}$. Secondly, qualitative findings such as *high IgG* as obtained by other labs should be allowed to add their share of evidence. Both

aspects could be integrated by replacing discrete findings with corresponding intervals, in the above examples $[1:4^{n-1}, 1:4^{n+1}]$ and, say, $[1:4096, 1:65536]$, and by adapting the reasoning method accordingly. Last but not least, the binary possibility distribution ONSET derives could be enhanced by a probability distribution assigning a chance of postconceptional infection to each individual case.

6.1. Comparison with symbolic approaches

Much of the work that the medical informatics community has contributed to assist with the interpretation of sparse clinical time series has been based on discrete logic. Only seldom reflected on is the fact that the discretization of analogue domains has its price. As a matter of fact, in most realistic applications involving temporal developments the combinatorial explosion in the number of cases to be considered boosts the complexity of the problem space to orders of magnitude far beyond human comprehension, as illustrated by the following estimate.

Let n be the number of sera drawn, c be the combinatorial number of possible outcomes of the different tests performed on one serum, and time be discrete with a resolution of one week. If a pregnancy lasts 40 weeks, the number of theoretically possible combinations is

$$c^n \binom{40}{n},$$

which amounts to 15,823,936,440 for 3 sera, 9 possible DT titers, and 13 possible IgM ISAGA indices (not regarding IgG avidity).

In an effort to reduce the size of the problem space cases with similar outcome are usually grouped together. To make the group or class specifications tractable, derived features are extracted from the raw data and numeric values are abstracted to qualitative terms or symbols (called *qualitative abstraction* in [Clancey 85]), allowing classificatory rules of the kind "if the titer is initially *low* and subsequently *rising*, infection is recent". However, strong points can be made that doing so is unduly simplistic, not only because of the problems with the divided difference touched on in Figure 6 b). Firstly, qualitative abstraction—depending on the choice of thresholds for the classification—is usually arbitrary. Secondly, as an untoward side effect of discretization, the diagnostic mapping becomes discontinuous allowing similar findings to be associated with disparate diagnoses; a fact that is hard to justify medically.

The latter point is illustrated by the graphical representation of the diagnostic criteria of IgG avidity depicted in Figure 8. The reader may verify that the graph is an exact graphical translation of the rules implicit in Table 2 and that determining the times of onset based on this graph is equivalent in effect to the application of the rules. For example, with an IgG avidity of 15% either method would restrict the time of onset of infection to be within the last three months, while with one of 16% it would be left entirely open; a discontinuity that is hardly natural. However, while the graphical representation can easily be adapted to reflect the continuity of natural courses of infection (as do the graphs of the other tests)[8], the table cannot; this is where the discrete approach loses.

Ironically, the diagnostic criteria of Table 2 were formulated only after a continuous monotonic, if not linear, dependency of IgG avidity on the age of infection had been observed [Lappalainen 93]. This reflects the predominant role of rules in communicating knowledge; yet, in the present case rules are clearly a weak mental prosthesis and their use is a detour leading to inferior results [Steimann 95b].

[8] The continuity property of ONSET's graphical approach is even better developped in its fuzzy version described in [Steimann 94d, Steimann 95a].

Despite these objections several laborartory decision-support systems relying on discrete symptom-disease mappings have successfully been deployed. For example, the HEPAXPERT-I system [Adlassnig 95] directly maps the discrete sample space of hepatitis A and B serology onto the set of possible diagnoses; the PRO.M.D. expert system shell for the interpretation of clinical chemistry test results, building on the symbolic programming language Prolog, adheres to the rule-based paradigm; it has been instantiated for disorders of lipid metabolism, syphilis serology, thyroid hormone diagnostics, etc. [Pohl 88].

Both HEPAXPERT-I and PRO.M.D. provide for unknown values, which is a common feature of rule-based systems. When dealing with sparsely sampled temporal developments such as the course of toxoplasma infection, unknown values are indeed at the heart of the problem. If, for example, the DT's value were known at all times, the time of onset of infection would be obvious (seroconversion!). Instead, however, it is not, and it is understood that a variable's value is unknown at all times but the sampled. Feeding unknown values to a dynamic system like ONSET is thus tautological. It is essentially the task of such a system to derive, if only approximately, values of variables otherwise unknown; indeed, by naming possible and impossible times of onset, ONSET constrains all serological variables' values at all times.

6.2. Alternative approaches

To account for temporal relationships, dynamic parameter models in the form of differential equations have been suggested (e.g. [Pohl 94]). Other researchers have investigated the impact of qualitative simulation on dynamic problems of the biomedical field, such as the monitoring of disturbances in the acid-base balance [Coiera 90, Ironi 90, Uckun 92]. In the pure qualitative approach actual parameter developments are matched against the predictions of qualitative models of alternative (patho)physiological hypotheses competing in the explanation of the monitored situation. It is questionable, however, if the problem dealt with in this article can at all be solved in qualitative terms: if not in the solution process, at least in its outcome definite references to the time line must be made, which is in contrast to the very spirit of qualitative simulation. Combined qualitative-quantitative models constraining qualitative simulation [Uckun 92] appear more adequate to tackle this problem.

The rather small number of samples (usually less than three tests performed on seldom more than three sera) renders several other symbolic approaches to the interpretation of clinical time series inept to solve the given problem, basically because their reasoning mechanisms require larger sample sizes. For example, temporal abstraction methods including the TOPAZ [Kahn 91], M-HTP [Larizza 92] and RÉSUMÉ [Shahar 93] systems are designed to handle the information overload encountered in long term patient monitoring. Their goal is to reduce large amounts of time-stamped data to high-level propositions assigned to temporal intervals. Serodiagnosis of recent toxoplasma infection in pregnancies, however, is the contrary problem: it has to make do with extremely few samples often lacking needed information and, in particular, leaving nothing to abstract.

Traditional numeric methods to identify temporal developments are mostly rooted in statistics and time series analysis: regression analysis, adaptive forecasting and Kalman filtering have successfully been applied to the interpretation of clinical time series [Allen 83, Gordon 86, Gordon 88, Avent 90, Challis 90, Sittig 90, Sittig 92]. However, like the temporal abstraction methods mentioned above these methods also rely on a minimum number of available samples. This is particularly untoward in their prospective employment where a diagnosis must be made after each sample including the first, a situation for which none of the statistical methods is designed.

7. Conclusion

We believe that the presented approach stands out with respect to its simplicity, naturalness and efficacy. In using continuous numeric specifications of courses of infection, in deriving temporal intervals of possible onsets and combining them through logical conjunction it is somewhat intermediate between conventional discrete (symbolic) and numeric methods. Like that of most other contemporary diagnostic support systems, its clinical relevance yet needs to be proven in practice.

8. Acknowledgements

The authors are indebted to Klaus-Peter Adlassnig and Arnold Pollak for their generous support.

Appendix: Proof

a) "$(1) \wedge (2) \Rightarrow (3)$"

$$d_x^-(t-t_\Omega) \le x_e'(t) \le d_x^+(t-t_\Omega)$$

$$\Rightarrow \int_{t_j}^{t_{j+1}} d_x^-(t-t_\Omega)dt \le \int_{t_j}^{t_{j+1}} x_e'(t)dt \le \int_{t_j}^{t_{j+1}} d_x^+(t-t_\Omega)dt$$

$$\Rightarrow \int_{t_j-t_\Omega}^{t_{j+1}-t_\Omega} d_x^-(t)dt \le x_e(t_{j+1}) - x_e(t_j) = x(p_i,t_{j+1}) - x(p_i,t_j) \le \int_{t_j-t_\Omega}^{t_{j+1}-t_\Omega} d_x^+(t)dt$$

b) "$(1) \wedge (3) \Rightarrow (2)$"

Surely there is an explaining course $x_e(t)$ satisfying (1). Now suppose this $x_e(t)$ does not satisfy (2). For each interval $[t_j, t_{j+1}]$ for which (2) is violated replace

$$x_e(t) := x(p_i,t_j) + D_x^-(t) + \frac{x(p_i,t_{j+1}) - x(p_i,t_j) - D_x^-(t_{j+1})}{D_x^+(t_{j+1}) - D_x^-(t_{j+1})}(D_x^+(t) - D_x^-(t)) \qquad (4)$$

where

$$D_x^-(t) := \int_{t_j}^{t} d_x^-(u)du \quad \text{and} \quad D_x^+(t) := \int_{t_j}^{t} d_x^+(u)du$$

Verify that if (3) holds (4) satisfies (2).

■

References

[Ades 91] AE Ades (1991) "Evaluating the sensitivity and predictive value of tests of recent infection: toxoplasmosis in pregnancy" *Epidemiol Infect* 107, 527–535

[Adlassnig 85] KP Adlassnig, W Horak (1995) "Development and retrospective evaluation of HEPAXPERT-I: a routinely-used expert system for interpretive analysis of hepatitis A and B serologic findings", *Artificial Intelligence in Medicine* 7, 1–24

[Allen 83] R Allen (1983) "Time series methods in the monitoring of intracranial pressure part I: problems, suggestions for a monitoring scheme and review of appropriate techniques", *J Biomed Eng* 5, 5–17

[Avent 90] RK Avent, JD Charlton (1990) "A critical review of trend-detection methodologies for biomedical monitoring systems", *Critical Reviews in Biomedical Engineering* 17(6), 621–659

[Aspöck 92] H Aspöck, A Pollak (1992) "Prevention of toxoplasmosis by serological screening of pregnant women in Austria", *Scand J Infect Dis: Supplement* 84, 32–38

[Bessières 92] MH Bessières, C Roques, A Berrebi, V Barre, M Cazaux, JP Séguéla (1992) "IgA antibody response during acquired and congenital toxoplasmosis", *J Clin Pathol* 45(7), 605–608

[Challis 90] RE Challis, RI Kitney (1990) "Biomedical signal processing (in four parts). Part 1: time domain methods", *Medical & Biological Engineering and Computing* 28, 509–524

[Clancey 85] WJ Clancey (1985) "Heuristic classification", *Artificial Intelligence* 27, 289–350

[Coiera 90] EW Coiera (1990) "Monitoring diseases with empirical and model-generated histories", *Artificial Intelligence in Medicine* 2, 135–147

[Desmonts 85] G Desmonts, J Couvreur: "Toxoplasmosis" in: RB Conn (ed): *Current Diagnosis* 7, W.B. Saunders Company, Philadelphia 1985

[Gordon 86] K Gordon (1986) "The multi-state Kalman Filter in medical monitoring", *Computer Methods and Programs in Biomedicine* 23, 147–154

[Gordon 88] K Gordon, AFM Smith: "Modelling and monitoring biomedical time series", in: JC Spall (ed) *Bayesian Analysis of Time Series and Dynamic Models*, Marcel Dekker, New York 1988

[Ironi 90] L Ironi, M Stefanelli, G Lanzola (1990) "Qualitative models in medical diagnosis", *Artificial Intelligence in Medicine* 2(2), 85–101

[Kahn 91] MG Kahn (1991) "Combining physiological models and symbolic methods to interpret time-varying patient data", *Methods of Information in Medicine* 30(3), 167–178

[Lappalainen 92] M Lappalainen, P Koskela, K Hedman, K Teramo, et al. (1992) "Incidence of primary toxoplasma infections during pregnancy in southern Finland: A prospective cohort study, *Scand J Infect Dis* 24, 97–104

[Lappalainen 93] M Lappalainen, P Koskela, M Koskiniemi, P Ämmälä, V Hiilesmaa, et al. (1993) "Toxoplasmosis acquired during pregnancy: improved serodiagnosis based on avidity of IgG", *Journal of Infectious Diseases* 167, 691–697

[Larizza 92] C Larizza, A Moglia, M Stefanelli (1992) "M-HTP: A system for monitoring heart transplant patients", *Artificial Intelligence in Medicine* 4, 111–126

[McCabe 83] RE McCabe, JS Remington (1983) "The diagnosis and treatment of toxoplasmosis", *Eur J Clin Microbiol* 2(2), 95–104

[Pohl 88] B Pohl, C Trendelenburg (1988) "PRO.M.D.–A diagnostic expert system shell for clinical chemistry test result interpretation", *Meth Inform Med* 27, 111–117

[Pohl 94] B Pohl, C Beringer, S Walther, J Melzer, F Burow, M Schmidt-Schauß, C Trendelenburg (1994) "Recent methods for knowledge-based interpretation systems with the PRO.M.D. expert system shell" (in German), *Laboratoriumsmedizin* 8, 577–582

[Remington 90] JS Remington, G Desmonts: "Toxoplasmosis", in: JS Remington, JO Klein, eds. *Infectious Diseases of the Fetus and Newborn Infant*, 3rd ed. (WB Saunders, Philadelphia 1990) 98–195

[Shahar 93] Y Shahar, MA Musen (1993) "RÉSUMÉ: A temporal-abstraction system for patient monitoring", *Computers and Biomedical Research* 26, 255–273

[Sittig 90] DF Sittig, M Factor (1990) "Physiologic trend detection and artifact rejection: a parallel implementation of a multi-state Kalman filtering algorithm", *Computer Methods and Programs in Biomedicine* 31, 1–10

[Sittig 92] DF Sittig, KH Cheung (1992) "A parallel implementation of a multi-state Kalman filtering algorithm to detect ECG arrhytmias", *International Journal of Clinical Monitoring and Computing* 9 , 13–22

[Steimann 94] F Steimann, M Hayde, B Panzenböck, KP Adlassnig, A Pollak (1994) "Fuzzy support for serodiagnosis: the ONSET program", *IEEE Engineering in Medicine and Biology Magazine* 13(5), 705–709

[Steimann 95a] F Steimann *Diagnostic Monitoring of Clinical Time Series* Doctoral dissertation, Technical University of Vienna, Austria 1995

[Steimann 95b] F Steimann: "A case against logic", in: *Proceedings of the MedInfo ´95*, (North-Holland, Canada 1995) 989–993

[Steimann 95c] F Steimann: "The interpretation of time-varying data with DIAMON-1", *Artificial Intelligence in Medicine* 8:3 (1996), in press

[Uckun 92] S Uckun, BM Dawant (1992) "Qualitative modeling as a paradigm for diagnosis and prediction in critical care environments", *Artificial Intelligence in Medicine* 4(2), 127–144