

Infectious agents

François Coutlée and Eduardo L. Franco

Summary

The detection and characterization of microbial agents in biological specimens are essential for the investigation of disease outbreaks, for epidemiologic studies of the clinical course of infections, and for the assessment of the role of infectious agents in chronic diseases. Methodological approaches depend on the infectious agent, the specimens analysed and the target populations. Although the diagnosis of infectious diseases has traditionally relied on direct microscopic examination of samples and on the cultivation of microbial agents *in vitro*, novel techniques with increased sensitivity and specificity are now being used on samples that can be more easily collected and transported to microbiology laboratories (e.g. dried blood spots on filter paper for nucleic

acid analysis). Direct detection techniques include the microscopic examination of specimens with special stains, antigen detection and nucleic acid detection by molecular assays. These assays are highly sensitive and provide rapid results for most agents. Genomic amplification assays greatly increase the sensitivity of nucleic acid-based tests by extensive amplification of specific nucleic acid sequences before detection. Real-time polymerase chain reaction (PCR) permits genomic amplification concurrently with detection of amplified products. Typing infectious agents requires additional investigation employing either serologic techniques to identify unique antigenic epitopes, or molecular techniques. These studies are important for

epidemiologic purposes, as well as for the investigation of pathogenesis, disease progression, and to establish causality between a disease and a microbial agent. Much of bacteriology has relied on growth of organisms on artificial media, and on identification of bacterial growth with biochemical, serological, or more recently, nucleic acid-based tests. The detection of specific antibodies from the host directed against pathogens is another strategy to identify current or past infections.

Introduction

The detection and characterization of microbial agents in biological fluids are important for the prevention, control, and management of infectious diseases in clinical

medicine, for the investigation of infectious outbreaks, for large-scale studies on the epidemiology of infectious agents, and for the assessment of the role of infectious agents in chronic diseases. Several approaches have been developed to attain these objectives depending on the nature of the infectious agent, the type of specimens available for analysis, the and populations evaluated (Table 10.1). Although the diagnosis of infectious diseases has traditionally relied on the direct microscopic examination of samples, and on the cultivation of microbial agents in *in vitro* systems, novel techniques with increased sensitivity and specificity are now being used on samples that can be more easily collected and transported to microbiology laboratories, such as the use of dried blood spots on filter paper for nucleic acid analysis. This chapter provides an overview of the approaches used to detect and identify infectious agents, investigate their relatedness, and characterize novel infectious agents in biological fluids. A comprehensive and detailed overview of available diagnostic molecular techniques by target agent is beyond the scope of this chapter, though the interested reader can find a wealth of detailed information on specific methods in several diagnostic microbiology and molecular epidemiology textbooks (1–3). Methods for the detection of microorganisms are classified into three categories: 1) direct detection techniques, 2) *in vitro* cultivation systems and 3) indirect detection based on serological methods that assess the host immune response against a putative infectious agent. This chapter also reviews the causal criteria for assessing the putative role of an infectious agent, and a chronic disease such as cancer, and the epidemiologic pitfalls due to measurement error inherent in

Table 10.1. Methods for the detection and analysis of microbial agents in biological fluids

1.	Direct detection of microbial agents
A.	Microscopic examination of specimens
B.	Microbial antigen detection
C.	Microbial nucleic acid detection
D.	Promising techniques: real-time PCR and matrix hybridization
2.	<i>In vitro</i> cultivation systems of microbial agents
A.	Non-cellular cultivation assays
B.	Cell culture systems
3.	Serological diagnosis of infectious diseases
A.	Screening assays
B.	Confirmatory assays

diagnostic testing for an infectious agent.

Direct detection techniques

Microscopic examination

Microorganisms are often directly detected in biological fluids by special stains, such as the Gram stain or acridine orange for bacteria; mycobacterial stains, based on the ability of mycobacteria to retain dyes after treatment with alcohol-acid decoloriser; nocardia stains; and calcofluor white for fungi. Wet mounts are used for detection of fungi or parasites. Potassium hydroxide is often added to better visualize yeast or hyphal structures. Enteric parasitic infections can be diagnosed by detection of ova in stools. Classically, the viral agents responsible for gastroenteritis are not detectable by cell culture, but usually are with electron microscopy. This is especially the case for the investigation of outbreaks caused by noroviruses. There are several limitations to these direct techniques. Although rapid and easy to perform, they are often insensitive and non-specific. Electron microscopy is a costly procedure

that requires the availability of an electron microscope and expertise in specimen processing and interpretation of results. Detection of a virus does not equate to active infection, as some individuals may simply shed a virus without active disease. Electron microscopy can, however, detect unsuspected or unknown agents (e.g. Severe Acute Respiratory Syndrome (SARS) coronavirus agent). Direct detection of microorganisms can also be accomplished histopathologically or cytologically by visualization of the pathogen itself with general-purpose stains, such as periodic acid-Schiff stain, or stains for substances produced by or contained in the pathogen, such as methenamine silver stain.

Detection of antigens from infectious agents

Simplified antigen detection assays are commonly used in diagnostic laboratories for a variety of microorganisms, including bacteria (e.g. group A streptococcus) or bacterial toxins (e.g. *Clostridium difficile*, enterohaemorrhagic *Escherichia coli*), viruses (e.g. varicella-zoster virus, influenza,

rotavirus), fungi (e.g. *Pneumocystis jirovecii*, *Cryptococcus neoformans*) and parasites (e.g. *Toxoplasma gondii*). Immunoassays involve the specific non-covalent binding of a microbial antigen to an antibody that is detected by a labelled ligand. Since antigen and antibody can react under a wide range of conditions, these assays can be applied in most biological fluids, including cerebrospinal fluid, stools, serum, respiratory secretions or urine. Several assay formats have been used (Table 10.2). The most widespread assays use antibodies that are fixed to a solid phase to separate unbound from bound antigens.

Direct or indirect immunofluorescence assays use an antiserum conjugated to a fluorochrome dye, fluorescein isothiocyanate. Some of these assays have been proven to reach sensitivity endpoints that are clinically relevant. Fluorescent staining assays are always subject to technical concerns and require expertise. The quality of specimens tested and sensitivity of the fluorescent assays are improved by including a cytocentrifugation step. The use of multiple antisera, directed against various infectious agents, is also an advantage of direct immunofluorescence (antisera pools).

Membrane EIAs have the advantage of being simple assays

that can be performed rapidly, do not require expertise or special equipment, and allow analysis of one specimen at a time. However, sensitivity is usually inferior to cell culture, and weak positive samples may be difficult to interpret. The same comments apply to latex agglutination techniques that have the added disadvantage of false-negative reactions due to prozone effects. The quality of these tests can be evaluated with periodic quality control panels.

The greatest drawback of the detection of microbial antigens is the limited sensitivity of these assays (e.g. *Legionella* fluorescent assays). Moreover, antigens can be degraded in clinical specimens, causing false-negative results. Intracellular antigens may not be detected as easily with these assays. Cross-reactivity, or high background staining in cellular material examined microscopically, can also generate false-positive results. The specificity of the antibody is influenced by the presence of non-microbial antigenic determinants co-purifying with microbial antigens during the steps of antigen production. Non-specific reactions can also be mediated by the antibody's F(c) portion, which can react with rheumatoid factor-like molecules in serum and some biological fluids. Finally, there may be cross-reactivity with other related organisms. Interestingly,

direct immunofluorescence assays have proven to be more sensitive than cell culture (see below) for some enveloped viruses that are susceptible to adverse transportation conditions. Microbial antigens can also be detected in biopsies using immunofluorescence or immunoperoxidase reagents.

Detection of nucleic acid from infectious agents

The limitations of traditional direct (see above) and cultivation methods (see below) have provided the impetus for the development and validation of new methodologies based on the detection and analysis of nucleic acids contained in biological samples. Nowadays, these assays are highly sensitive, provide rapid results for most microbial agents, and can be partially automated. The unique sequence specificity of DNA from a given species permits the design of assays that are highly specific to a target agent. Performance of the test is not affected by death of the organism due to antimicrobial therapy, or to transportation or storage of specimens under suboptimal conditions (provided that extreme conditions that can affect DNA integrity can be avoided). DNA is one of the most stable and chemically resistant biological molecules in nature. Intracellular

Table 10.2. Most common assay formats for detection of microbial antigens in biological fluids

Assay Format	Time for completion	Potential for automation	Control of cellularity
Microtiterplate EIA	<4 hours	+	-
Particle agglutination	15 minutes	-	-
Membrane immunoassays	15 minutes	-	-
Direct or indirect IF	30-60 minutes	-	+

EIA, enzyme immunoassay; IF, immunofluorescence.

organisms can be detected after a DNA extraction treatment of samples. These techniques can also detect organisms involved in latent infections. Nucleic acid-based assays have been able to demonstrate integration into the human genome of viral DNA (e.g. hepatitis B virus (HBV) and human papillomavirus (HPV)). These assays can be quantitative or detect microbial mRNA, permitting the analysis of transcriptional activity of a microbial agent, such as for cytomegalovirus. Recombinant DNA technologies, as well as oligonucleotide synthesis strategies, have facilitated the synthesis of large quantities of reagents that can be standardized and quality-controlled more easily than for other diagnostic modalities. The cost of production of reagents has decreased substantially in recent years. Similar protocols for reagent synthesis can be used for different agents, since DNA is the target for all assays irrespective of the agents detected. This is in contrast to immunoassays, for which antibodies are produced by immunization of animals for polyclonal antibodies, and by hybridoma formation for monoclonal antibodies.

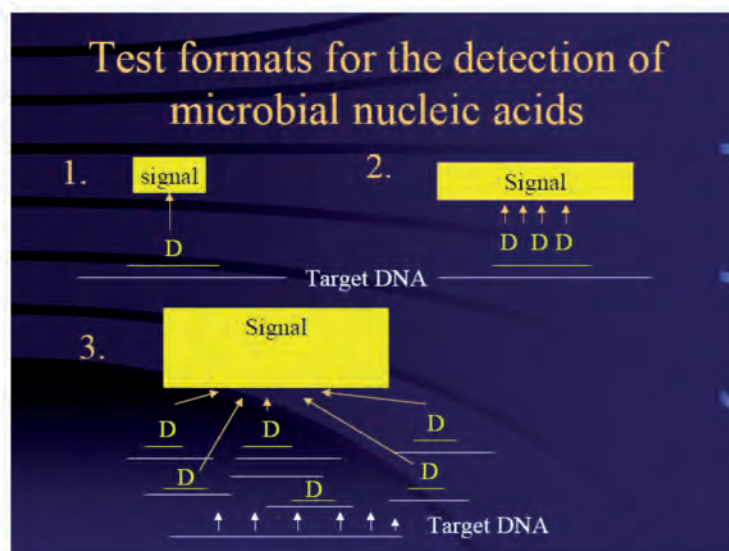
Although the use of direct nucleic acid assays was initially impeded by a lack of sensitivity, since the 1990s genomic amplification techniques have revolutionized diagnostic microbiology. The reliance on radiolabelled probes, initially to detect specific nucleic acid sequences (Figure 10.1), was also an important limitation for diagnostic laboratories. Radioactive probes had a short functional half-life and sometimes necessitated prolonged exposure to photographic plates to reach adequate sensitivity. The expense of the facilities required to manipulate radioactive material and dispose of it were significant drawbacks.

Fortunately, non-radioactive labels (i.e. enzyme-labelled probes, avidin-biotin systems, Europium, acridinium esters and others) have been successfully used with these technologies. Assays manufactured by Gen-Probe, with acridinium-labelled single-stranded DNA probes, have been used to detect *Chlamydia*, mycobacterial, fungal or Neisseriaceae rRNA by chemiluminescence. They are now mostly used to identify dimorphic fungi or mycobacterial isolates isolated by culture. Another format of nucleic acid detection tests, the signal amplification assays (branched DNA tests and Hybrid capture), have significantly increased the endpoint sensitivity of DNA-based assays (Figure 10.1). In these assays, the signal is amplified without amplification of the nucleic acid target (see below), thus avoiding carryover contamination and false-positive results. These assays are often less sensitive analytically than genomic amplification techniques, but reach

useful clinical sensitivity endpoints (e.g. for human immunodeficiency virus (HIV), hepatitis C virus (HCV) or HPV), and are also highly specific.

Genomic amplification assays have greatly increased the sensitivity of nucleic acid-based tests by extensive amplification of the target nucleic acid sequence before detection (Figure 10.1). Several amplification technologies have been devised, as described in Table 10.3. Of note are those based on thermal cycling amplification, such as the polymerase chain reaction (PCR) (Roche Molecular Systems), and those based on isothermic amplification, such as transcription-mediated amplification (TMA) (Gen-Probe), Nucleic Acid Sequence-Based Amplification (NASBA) (Organon Technika) or strand displacement amplification (SDA) (Becton-Dickinson). PCR is the amplification format that has been the most widely used to develop assays available commercially and as in-house assays. When appropriately optimized and

Figure 10.1. Assay formats for the detection of nucleic acids in biological fluids. 1. Direct detection of nucleic acid (endpoint analytical sensitivity $\pm 10^5$ copies per test). 2. Signal amplification tests (endpoint analytical sensitivity $\pm 10^3$ copies per test). 3. Genomic amplification techniques (endpoint analytical sensitivity ± 1 copy per test)



validated, they have consistently proven to be highly sensitive and specific. They can detect pathogens present in low quantities that are slow-growing or cannot be cultivated, and even infectious agents not yet discovered. Multiplex PCR assays can simultaneously detect several pathogens. By adding a reverse transcription step, RNA viruses can be detected with these molecular techniques. The complexity of some viral families (e.g. enteroviruses, HPVs) requires the use of consensus amplification assays to detect all relevant genotypes. These techniques can also be quantitative (Table 10.3). Measures of microbial loads are

important information that can be predictive of existing disease, for deciding on initiation of treatment, or assisting the follow-up of treated individuals to assess response or resistance to therapy. For some pathogens, such as mycobacteria, PCR has been used to complement cultivation methods.

The exquisite sensitivity of amplification assays can cause problems in less experienced laboratories. Contamination of reagents by carryover of previously synthesized amplicons can generate false-positive results, but these mishaps can be prevented (Table 10.4). Good laboratory practices, and the use of separate working

zones and plugged micropipette tips, effectively curtail the risk of contamination. These techniques are now widely employed without problem in accredited diagnostic laboratories. One limiting step of these assays is the extensive extraction procedures that are sometimes required to analyse samples. Automated extraction instruments resolve this issue in well-equipped laboratories. Finally, inhibitor substances that impede the amplification process can generate false-negative results, but can be screened for by the use of internal controls or amplification of human genes to assess specimen quality.

Table 10.3. Assay formats for amplification of nucleic acids

Assay Format	Manufacturer	Cycling temperature	Qualitative detection	Quantitation
PCR	Roche Molecular Systems	thermal cycling	yes	yes
TMA	Gen-Probe	isothermic	yes	yes
NASBA	Organon Technika/ BioMérieux	isothermic	yes	yes
SDA	Becton-Dickinson	isothermic	yes	no

PCR, polymerase chain reaction; TMA, transcription-mediated amplification; NASBA, nucleic acid sequence-based amplification; SDA, strand displacement amplification

Table 10.4. Most useful procedures to control false-positive results in nucleic acid amplification assays due to contamination

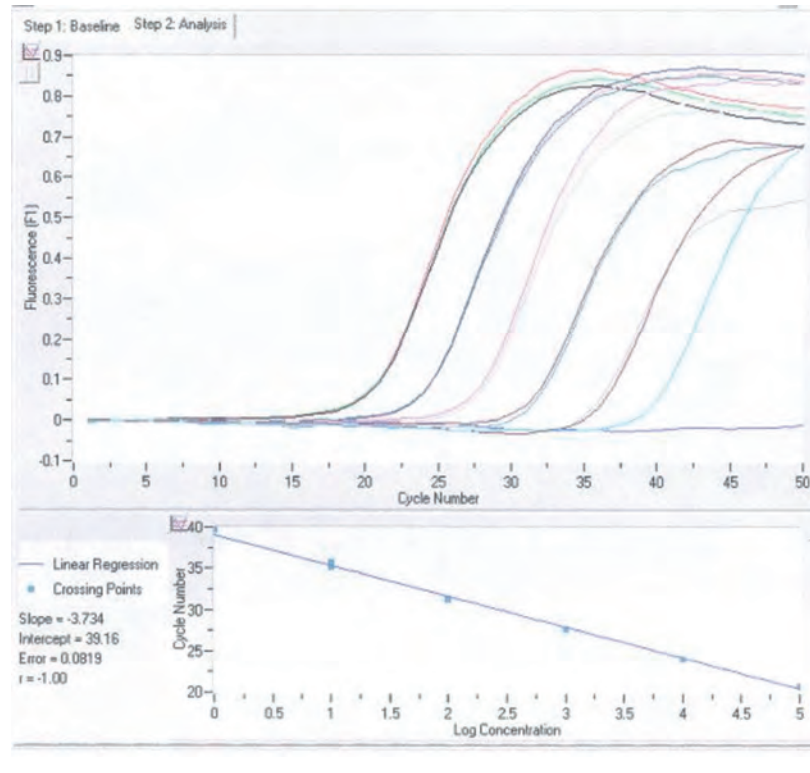
Good laboratory practices (supported by on-site manuals describing standard operating procedures) to prepare samples at perform testing
Separated working zones (pre- and post-PCR areas)
Dedicated instruments for each working zone
Aliquot reagents to avoid repeated use of reagents, especially the master mix
Use of plugged micropipette tips
Avoidance of strongly reactive positive controls
Chemical or enzymatic destruction of contaminating amplicons

New trends in nucleic acid detection tests

Real-time PCR is a new development in the science of genomic amplification. Amplification is performed concurrently with detection in a closed tube, significantly reducing the time to complete the assay and potential for cross-contamination by carryover. Probes are labelled with various fluorophors, and multiple targets can thus be detected. Quantitation of targets is typically done in the logarithmic phase of amplification, which provides for more reproducible measurements of the analyte. The detection of a positive signal is obtained during amplification, as shown in Figure 10.2, significantly shortening the time required to complete testing.

In the future, the use of microarrays will allow the detection of panels of infectious agents that will be selected depending on the disease screened and sample tested. These assays are based on the attachment to solid supports of up to thousands of oligonucleotide probes, generating a matrix of probes. Labelled amplified products are then hybridized with these fixed probes, and the specific signals generated by fixed labelled amplicons can be detected and analysed with a computer. One system utilizes photo-activation for the chemical synthesis of small DNA fragments directly onto a silicate solid support (silicon chip methodology), generating complex arrays of probes. After hybridization of labelled amplicons with these arrays, the silicon surface is screened with a scanning laser confocal fluorescence microscope. These assays can analyse complex mixtures of nucleic acids. The first systems developed using the chip technology successfully analysed HIV resistance to antiretroviral

Figure 10.2. Amplification and detection of human DNA by real-time PCR. A titration curve of human DNA was tested and results were plotted in a titration curve (shown above). The rising curves indicate the presence of DNA amplified and detected in the assay. This information is provided online during the performance of the assay. The triplicates for each dilution show excellent reproducibility with curves being almost superimposed. The regression line obtained with these dilutions was excellent, with $r = -1.00$



treatment. There are several drawbacks to these techniques, including the complexity of fabrication of the probe arrays, instruments required to perform these tests, and cost. These assays will still require extensive validation before application on populations or cohorts of individuals. These promising techniques are under investigation and could be applied in diagnostic and molecular epidemiologic laboratories in the next decade.

Molecular techniques for genotyping

Typing infectious agents requires additional investigations, which

either employ serologic techniques to identify unique antigenic epitopes, or molecular techniques to analyse the microbial genome. These studies are important for epidemiologic purposes, and also for the investigation of pathogenesis, disease progression and the causal association between a disease and a microbial agent. Amplification methods are now the cornerstone for the molecular component of the epidemiologic investigation of infectious diseases, and are replacing phenotypic techniques (e.g. biotyping, susceptibility testing, serotyping, bacteriophage typing, and multilocus enzyme electrophoresis). A variety of DNA-based methods can be used to

study the relatedness of different isolates of a species. Non-genomic amplification methods include: bacterial plasmid analyses, restriction endonuclease analysis of bacterial DNA or Southern blot analysis of restriction fragment length polymorphisms (RFLP), and pulsed-field gel electrophoresis (PGFE) of chromosomal DNA. The latter technique includes ribotyping for bacteria. Ribosomal sequences are highly conserved and could react with a wide range of bacterial species. All bacteria carry the ribosomal operons and are thus typeable. Ribotypes are stable, which facilitates the investigation of outbreaks. Variable regions of the microbial genome are ideal targets for these analyses. The study of insertion sequences (e.g. IS6110 DNA sequence for *Mycobacterium tuberculosis*) can also be used to investigate laboratory cross-contamination, identify sources of infection in outbreaks, and assess if a new recurrence is due to the initial organism or to reinfection, or if an infection is caused by multiple isolates.

Several genotyping methods have been adapted to PCR. The amplification step obviates the need for isolating the agents in culture and can be applied directly on samples. PCR-RFLP involves the digestion of PCR-generated amplicons with restriction enzymes, and depending on the various restriction patterns obtained, polymorphism can be studied. This low-cost technique is simple, easy to perform, and can accommodate testing of a large number of samples rapidly. However, only a limited number of DNA sites are analysed. PCR-single stranded conformation polymorphism (SSCP) is a technique in which radiolabelled amplicons are denatured and migrated in a non-denaturing polyacrylamide gel.

The conformation of the migrating DNA strand is dependent on the nucleotide sequence of the amplicon, which ultimately affects the migration pattern of the latter. Single nucleotide changes can be detected with this technique. It has the advantage of analysing the complete amplicon, in contrast to PCR-RFLP, but it requires manipulation of radioactive reagents, it is more time-consuming, especially to optimize migration conditions, and it may miss some polymorphisms. Arbitrarily-primed PCR (AP-PCR), or randomly amplified polymorphic DNA (RAPD), is based on the observation that short non-specific primers of 10 nucleotides will hybridize and amplify random DNA sections of chromosomes that differ between genotypes. Since the number and locations of binding sites of short primers will vary, differences between genotypes can be established. Identification of suitable primers may require considerable effort. The technique has been described mainly for bacteria and fungi.

The heteroduplex mobility assay (HMA) is based on the hybridization of PCR amplicons generated from different isolates of a microbial agent. Duplexes containing bulges because of mismatches between amplicon strands from different genotypes will migrate differently during electrophoresis in neutral polyacrylamide gels. The relative retardation of migration is proportional to the DNA distance between genotypes analysed. This simple and rapid technique is limited by its capacity to detect genetic differences of at least 2%. The usefulness of this method was demonstrated in the analysis of viral quasi-species (as for HIV and HCV).

Automated sequencing facilities represent a significant, important improvement in nucleic

acid-based tests for genotyping. PCR sequencing determines the nucleotide sequence of microbial DNA, thus permitting identification of the implicated microorganism. It is also essential for phylogenetic analysis and very useful for molecular epidemiology purposes, such as in the investigation of outbreaks or for examining the possible causal role of infectious agents in diseases. It is considered the gold standard method for genotyping. Although still a costly procedure, this has become less of a problem in recent years due to the availability of more affordable instrumentation. Results for each nucleotide position are generated. PCR sequencing does not require knowledge of the pathogen's complete DNA sequence. However, it does generate important quantities of data that must be systematically analysed. The analysis of RNA genomes is further complicated by the existence of several viral species of quasi-species, which add complexity to the genotyping process.

***In vitro* cultivation systems**

Non-cellular cultivation assays

Much of diagnostic microbiology, especially bacteriology, has relied on the growth of organisms on artificial media. Bacterial growth has been identified using biochemical methods with antisera more often using agglutination tests (e.g. latex for β -haemolytic streptococci or whole-organism suspension for *Salmonella* and *Shigella* serogroups), or by nucleic acid-based tests. For most bacteria, cultivation on artificial media is the mainstay of diagnostic microbiology. Enriched all-purpose media, such as blood or chocolate agar, are used to grow common human pathogens. Selective media

can also be used to screen for pathogens in the presence of normal microbial flora. Subculture in broth media increases the sensitivity of culture, but decreases its specificity; however, the microorganism is isolated and can be analysed more easily. Antimicrobial susceptibility testing can also be performed and used as an epidemiologic marker on isolated bacterial or fungal isolates. Cultivation techniques are often less sensitive for fastidious organisms, or when patients have started antimicrobial therapy before specimens were obtained for culture. Prolonged periods of incubation may be required for pathogens that grow slowly, such as several mycobacterial species and fungal dimorphic agents. Unfortunately, some key pathogenic bacteria cannot be readily cultivated *in vitro*, such as *Treponema pallidum*, *Mycobacterium leprae*, *Bartonella henselae* and *Tropheryma whippelii*. Recovery of bacterial pathogens in some specimens may be impeded by abundant normal bacterial flora competing for nutrients contained in artificial media. Moreover, pathogens may have similar phenotypes as bacterial agents from the normal flora. For instance, enterotoxin-producing strains of *Escherichia coli* (*E. coli*) that cause diarrhoea are undistinguishable from non-virulent *E. coli* strains.

Cell culture systems

Cell culture allows the detection of a wide range of viruses and the presence of mixed viral pathogens in specimens. After adding a specimen to a monolayer of cells obtained *in vitro*, the presence of a virus in cell cultures can be detected by the distinctive cytopathic effect on cells caused by viral replication (e.g. herpesviruses), by haemadsorption or haemagglutination (e.g. influenza

viruses), or with virus-specific fluorescein-labelled antisera (e.g. cytomegalovirus). Viral isolates can be further characterized by molecular techniques for genotyping, antiviral susceptibility testing or immunoreagents for serotyping. However, the requirement for maintenance of several cell lines to support growth of most human viral agents limits cell culture to specialized laboratories. Moreover, propagation of some viruses, such as HIV, represents a significant biohazard for laboratory workers and requires Level-3 containment facilities. Likewise, some cell lines, such as Vero cells, can support the growth of the SARS agent, which represents a considerable biohazard for technologists. The viability of fragile viruses, mostly enveloped viruses, is adversely affected by inadequate transportation and storage conditions. For instance, the rate of positive cultures is lower in summer than in winter months for herpes simplex viruses. Also, Varicella-zoster virus is more frequently detected by direct immunofluorescent tests on samples than by cell culture. Some fastidious viruses do not grow well in cell culture. Furthermore, cell lines are not available for many key human pathogens, including rotavirus, norovirus, hepatitis A virus, HBV, HCV and Epstein-Barr virus (EBV). The delay before a cell culture turns positive is also a limitation of this procedure, as traditionally cell cultures are kept for 7 to 28 days. Shell vial spin amplification, most commonly used for cytomegalovirus and respiratory viruses but also for some fastidious bacteria (e.g. *Bartonella henselae* or *Francisella tularensis*), shortens this delay. In this procedure, specimens are added to a cell culture monolayer in a vial, centrifuged at low speed after inoculation, incubated, and

reacted with a fluorescent antibody against viral antigens associated with a replicating virus. Detection of viral agents thus becomes possible before the development of a cytopathic effect.

Indirect detection via serological methods

The detection of specific host antibodies directed against pathogens is another strategy used to identify current or past infections. The detection of antibodies against infectious agents can be performed in serum, as well as cerebrospinal fluid (e.g. arboviruses). The diagnosis of acute infection is usually based on a four-fold increase, or more, of specific antibody titres in paired acute and convalescent sera obtained at two- to four-week intervals (e.g. respiratory viruses), or by the presence of specific immunoglobulin M (IgM) antibodies (e.g. Human parvovirus B19, *Toxoplasma gondii*). Detection of IgM antibodies is less sensitive in immunosuppressed individuals or newborns, however, and it can also be affected by heterologous responses and interference with rheumatoid factor-like molecules in the serum. Direct detection or cultivation methods provide faster results for diagnostic purposes. Chronic infections can be diagnosed by testing a serum to detect specific immunoglobulin G (IgG) antibodies against a preparation of the agent's antigens, such as HCV serology, or a panel of IgG antibodies directed against various microbial antigens that indicate current or resolved infection (HBV or EBV). For example, serologic methods are used for the diagnosis of acute primary EBV infection (by detecting IgG and IgM antibodies against viral capsid antigen, early antigen and Epstein-Barr nuclear antigen), and

for the screening of EBV-associated nasopharyngeal carcinoma (mainly immunoglobulin A (IgA) against early and capsid antigens), while molecular techniques are most useful for the diagnosis of EBV-related lymphomas. Serology testing is very valuable for the diagnosis of chronic infections, such as HIV or viral hepatitis. In contrast with direct detection methods, serological testing provides information on past and current infection status of an infectious agent. Serological assays are also adequate tests to evaluate response to vaccination. Serology is most frequently used for the diagnosis of viral infections, but can be valuable in identifying individuals infected with protozoan or metazoan parasites and fungi.

Several techniques have been used to detect antibodies directed against infectious agents, including complement fixation, immunodiffusion, particle (e.g. latex) or erythrocyte agglutination, immunofluorescence, and enzyme immunoassays (also known as enzyme-linked immunosorbent assays (ELISAs)). In a diagnostic microbiology laboratory, most of today's serologic tests are performed with commercially available EIA or immunofluorescence formats. For several agents, screening is performed with EIA tests because of the ease, rapidity and low cost of this assay format. Better purification of viral antigens has resulted in improved sensitivity and specificity of EIAs for viral hepatitis diagnosis. Improved assays were thus designated as second- and third-generation assays. Positive results are then confirmed by a more specific technique that is often more cumbersome and costly. These techniques include recombinant immunoblot assays, radioimmunoprecipitation assays or Western blot assays (e.g. HIV and HCV).

Implicating infections as causes of cancer and other chronic diseases

The operational epidemiologic definition of a cause is a factor that alters the risk of disease occurrence. For infectious diseases, the definition has been more mechanistic: a cause is either a factor that must exist for disease to occur (i.e. is necessary) or always produces disease (i.e. is sufficient). A microbial agent is a necessary, and sometimes sufficient, cause of an infectious disease, depending on the interplay between agent, host and environmental factors. On the other hand, the situation is less clear for cancer: a group of diseases of multifactorial etiology, which ultimately result from the interaction between external environmental causes and the internal genetic makeup of the individual. Few of the accepted causes of human cancer are deemed necessary (e.g. HPV infection in cervical cancer) or sufficient (e.g. possibly some of the high penetrance cancer genes). Unlike most infectious diseases, cancer has a long latency period, which underscores the succession of time-dependent events that are necessary for normal tissue to develop into a malignant lesion and ultimately progress into invasive cancer. Carcinogenesis is a multistage process where final onset of disease is a function of the combined probabilities of relatively rare events occurring in each stage. These events depend on a myriad of factors related to carcinogen absorption and delivery to target cells, metabolic activation, binding with relevant gatekeeper or caretaker genes, and to the ability of the affected tissue to reverse these initiating processes. Also to be considered is the contribution of promoters, which will favour

cell proliferation with consequent selection of clones with selective growth advantage within the surrounding tissue. Eventually, other factors will facilitate progression of a precancerous lesion to invasive cancer, and thus also contribute a causal role in carcinogenesis.

Historically, causal relationships in infectious diseases have been assessed using the mechanistically based Henle-Koch's postulates, which are based on the expectation that the microbial agent must be necessary, specific, and sufficient for the disease to occur. These postulates are only of indirect help in assessing cancer or chronic disease etiology, since they imply the causation of the immediate infectious disease or condition that originated from the agent, and not the final malignant process at the end of a lengthy chain of events triggered by the infection itself. A case-in-point is the causal pathway represented by the acquisition of infection with HBV in non-immune individuals, followed by the development of acute hepatitis, chronic hepatitis, and finally, many years later, the onset of hepatocellular carcinoma. Each step in succession affects smaller proportions of patients than the previous. Henle-Koch's postulates are useful up to the first or second steps of this pathway, but are of no guidance for the imputation of a causal link between the pathways' beginning and terminal events.

The reasoning into what constitutes the criteria for judging whether or not a given risk factor is a cause of cancer has primarily evolved from the so-called Bradford Hill criteria (4), a subset of which are referred to as the Surgeon General's criteria (5). These criteria were first proposed at the time of a vigorously debated health issue of the early 1960s, namely the interpretation of the accrued evidence on the role

of tobacco smoking as a cause of lung cancer. Hill's nine criteria were: strength of the association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experimental evidence and analogy. In his seminal paper (4), he downplayed the importance of specificity, plausibility and analogy, which are viewed today as non-essential and can even be considered counter-productive distractions to the discussion of any possible cause-effect relationship in cancer. Unfortunately, however, he also concluded that "...none of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*" (4). If published today, the second part of that statement would have been disputed immediately. Clearly, temporality is a necessary causal criterion, and biological gradient, consistency, and strength of the association are among the most frequently used in cancer risk assessment (reviewed in (6)).

Although highly persuasive in establishing causality, the availability of experimental evidence from randomized controlled trials is more the exception than the rule in public health. In the case of an infectious cause of cancer, one may include the results from vaccine trials of HPV and HBV, as well as post-deployment surveillance of these vaccines in different populations, which have provided strong evidence that these agents are unequivocally causal regarding their respective malignant diseases (i.e. cervical neoplasia) (7,8). (As of this writing, trials have shown a reduction in risk of precancerous lesions only, and not yet of cervical cancer and hepatocellular carcinoma (9)). Typically, the change in prevalence of a disease is observed after the prevalence of a causal determinant

has been modified, subsequent to allowing for sufficient latency. More often, epidemiologists derive evidence from observational studies, such as case-control and cohort studies, which are prone to biases in interpretation because of confounding, measurement error (see below), and other issues that preclude isolating the effect of a single factor on causation.

Although useful for environmental, occupational and lifestyle determinants, Hill's criteria do not capture very well the evidential foundation of causal claims for microbial agents and their respective malignant diseases. Fortunately, useful guidelines for causal attributions involving infectious agents have been proposed (10–12). Summarized in Table 10.5, they are correlated with the original criteria formulated by Hill. These causal criteria take into account the knowledge about the timing, specificity and level of immune response against putative viruses, or the advances in nucleic acid detection methodology as used in modern molecular epidemiologic investigations.

In summary, what prevails today is an operational definition of cause, which incorporates the criteria required in different settings. Determining an exposure and intermediate endpoints related to an infectious agent depends on the type of mechanism being studied and its particular set of circumstances (13). Decisions concerning the etiologic role of specific infectious exposures must be a dynamic process that entertains both scientific and public health issues, and is constantly updated as new knowledge from more insightful and valid epidemiologic studies becomes available.

Epidemiologic pitfalls due to measurement error

Epidemiologic common sense has it that improper ascertainment of exposure variables will bias the relative risk (RR) estimates, generally towards the null hypothesis, if the misclassification is random and nondifferential with respect to the outcome (being a case of the disease or not). If the measurement error is not random or nondifferential with respect to the outcome, the direction and degree of the bias are difficult to predict. Although modern molecular methods to determine exposure to infectious agents have attained a substantial degree of accuracy, errors related to sampling, variations in viral load, and other mishaps all contribute to exposure misclassification. The following paragraphs describe the effects of misclassification in specific circumstances typical of epidemiologic studies, which attempt to examine the putative causal role of an infectious agent for a chronic disease such as cancer. In particular, the impact of measurement error on the prevalence of infection in field surveys, and on the magnitude of the association between the infectious agent and cancer in epidemiologic studies, is illustrated. Both of these issues are germane to our interpretation of the putative causal role of an infectious agent in a chronic disease that follows the exposure after a long latency period. (See Chapter 8 for additional discussion of misclassification and measurement error.)

Table 10.5. Criteria or guidelines used in attributing causality to candidate infectious, lifestyle and environmental risk factors* in the genesis of cancer and other chronic diseases of long latency

	Role of Environment & Lifestyle		Role of Infections	
	Hill (1965)	Evans (1976)	Evans and Mueller (1990)	Fredricks and Relman (1996)
Strength of the association: magnitude of the relative risk for the factor and incident disease or mortality			Presence of viral marker should be higher in cases than in controls Incidence of tumour should be higher in those with the viral marker than in those without it Geographic distributions of viral infection and tumour should coincide	Nucleic acid from agent should be present in most cases and preferentially in affected organs
Consistency: findings are replicated by studies in different populations		IgG and IgM antibodies to the agent regularly appear during illness		Molecular or serological evidence should be reproducible
Specificity: Exposure to factor tends to be associated with only one outcome		Antibody to no other agent should be similarly associated with the disease unless a cofactor in its production		Few or no copy numbers should occur in hosts or tissues without disease
Temporality: Onset of exposure should precede with sufficient latency to disease occurrence		Antibody to the agent is regularly absent prior to the disease and exposure to the agent	Appearance of viral marker should precede the tumour	Detection of DNA sequence should predate disease
Biological gradient: Dose-response relation between factor and rate of disease development				Copy number should decrease or become undetectable with disease regression (opposite with relapse or progression)
Plausibility: Does the association make sense in relation to the existing knowledge of likely mechanisms that could be affected by exposure?		Absence of antibody to the agent predicts susceptibility to both infection and the disease produced by the agent Antibody to the agent predicts immunity to the disease associated with infection by the agent		Tissue-sequence correlates should be sought at the cellular level using <i>in situ</i> hybridization
Coherence: Does the association conflict with other knowledge about the disease?				
Experimental evidence: Data from randomized controlled trials that eliminate exposure or its consequences			Immunization with the virus should decrease the subsequent incidence of the tumour	
Analogy: Is there a comparable exposure-disease association that seems analogous?				Microorganism inferred by DNA sequence should be consistent with the biological characteristics of that group of organisms

* Or their respective exposure circumstances.

Bias in prevalence surveys

The effect of misclassification on the presumed prevalence of an infectious agent can be understood if the diagnostic performance of the chosen laboratory test is known, particularly its sensitivity (S) and specificity (W) with respect to the true exposure or infection status. The formula (14) to correct for the bias is as follows:

$$P_c = (P_u + W - 1) / (S + W - 1)$$

where P_c and P_u are the corrected and uncorrected prevalence rates, respectively.

Depending on the true prevalence rate that must be estimated via the test and its diagnostic performance, the estimated rate can be a gross overestimation of the true prevalence rate. For instance, for a rare infectious exposure prevalent among 2.5% of the individuals in the target population, a test with false-negative and false-positive rates of 10% ($S = W = 90\%$) and 20% ($S = W = 80\%$) will be positive 12% and 21.5% of the time in the survey, respectively, thus substantially overestimating the true rate. Under such conditions, the bias always results in overestimation of the prevalence rate and is more influenced by the specificity than by the sensitivity of the assay. Lowering sensitivity has only a moderate biasing effect on the presumed rate.

Bias in the magnitude of the association

As above, if the diagnostic properties of the assay that were used to ascertain exposure to an infectious agent are known, one can correct the estimated measure of the association for the relation between agent and disease. For instance, in a case-control or cross-sectional study, the formula (14) for correcting the odds ratio (OR) is as follows:

$$OR = \frac{(W_1 n_1 - b)(S_2 n_2 - c)}{(W_2 n_2 - d)(S_1 n_1 - a)}$$

where S = sensitivity, W = specificity, n is the number of subjects, and the subscripts 1 and 2 indicate that the information is for cases or controls, respectively. The frequencies a , b , c , and d are the study's 2x2 table frequencies as follows: a = exposed cases, b = unexposed cases, c = exposed controls and d = unexposed controls.

It is possible to simulate the impact of measurement error of an infectious exposure that causes a precursor cancerous lesion, affecting 2.5% of the population after a specified period of time (e.g. high grade cervical intraepithelial neoplasia (HGCIN)) (15). For illustration, assume that the prevalence of the putative agent (i.e. HPV) is 20%, and the underlying RR for the relation with the lesion outcome is 100. Under conditions of perfect measurement of lesion outcome (HGCIN and non-HGCIN), increasing misclassification of HPV status leads to biased estimates of RRs towards unity. For instance, at 10% misclassification ($S = W = 90\%$), the original RR of 100 is erroneously measured as $RR = 19$. At 30% misclassification, the bias is so severe that the measured RR is just below 4.

In practice, study validity is further aggravated by concomitant misclassification of the outcome, which is a real concern in cohort studies, as for ethical and practical reasons they may have to rely on pre-invasive lesions as endpoints. On the other hand, case-control studies of invasive cancer are far less likely to be affected by outcome misclassification, but are prone to differential exposure misclassification, as detection of the infectious exposure may vary

between cases and controls. In the case of HPV infection and cervical cancer, detection of the former is done in exfoliated cervical cells, which results in sampling differences between cases and controls. Moreover, the effects of fluctuation in viral load, transience of HPV infection, and other factors inherent to the dynamics of the infection make single testing for a virus, such as HPV, less likely to represent past exposure for controls than for invasive cancer cases. Capturing the actual exposure experience to HPV that led to cancer would have required sampling the cases' cervix earlier, when the infection was at a comparable state to that of the controls. The biasing effects of these two errors are in the same positive direction away from the null hypothesis (i.e. they produce RRs that are higher than the one truly underlying the relation between HPV and cervical cancer in the same population).

There is one important source of misclassification that cannot be corrected by knowledge of test parameters: it is caused by the biological variation in the ability to detect exposure to the agent over time. Again, the HPV-cervical cancer example is illustrative. Most instances of HPV infection are transient. It is clear, therefore, that collection of a single cervical specimen at the time of enrolment in a cohort study, or at the time of diagnosis of HGCIN, or of invasive cervical cancer in a case-control study, provides little assurance that the laboratory determination of the HPV positivity of that specimen accurately reflects the relevant past exposure to HPV infection that the subject may have had. Infections with low viral load may be labelled erroneously as HPV-negative. A subject with a mildly productive transient infection at the

time of testing may be classified as HPV-positive in epidemiologic studies based on single-specimen assessment of exposure, regardless of whether the design is cohort or case-control. Such studies will

also attribute exposure status to false-positive specimens resulting from contamination. The latter subjects' non-exposed status can be ascertained with greater validity if one determines a cumulative

exposure status based on detection of HPV in multiple specimens collected over time in repeated measurement studies (16).

References

1. Murray PR, Baron EJ, Jorgensen JH *et al.*, editors. *Manual of clinical microbiology*. 9th ed. Washington (DC): ASM Press; 2007.
2. Mahon CR, Lehman DC, Manuseelis G. *Textbook of diagnostic microbiology*. 3rd ed. St. Louis (MO): Saunders-Elsevier; 2007.
3. Riley LW. *Molecular epidemiology of infectious diseases: principles and practices*. Washington (DC): ASM Press; 2004.
4. Hill AB (1965). The environment and disease: association or causation? *Proc R Soc Med*, 58:295–300. PMID:14283879
5. Surgeon General's Advisory Committee on Smoking and Health. *Smoking and health*. Washington (D.C.): U.S. Department of Health, Education and Welfare, Public Health Service, PHS Publication no. 1103; 1964.
6. Weed DL, Gorelic LS (1996). The practice of causal inference in cancer epidemiology. *Cancer Epidemiol Biomarkers Prev*, 5:303–311. PMID:8722223
7. Harper DM, Franco EL, Wheeler CM *et al.*; HPV Vaccine Study group (2006). Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial. *Lancet*, 367:1247–1255. doi:10.1016/S0140-6736(06)68439-0 PMID:16631880
8. FUTURE II Study Group (2007). Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med*, 356:1915–1927. doi:10.1056/NEJMoa061741 PMID:17494925
9. Montesano R (2002). Hepatitis B immunization and hepatocellular carcinoma: The Gambia Hepatitis Intervention Study. *J Med Virol*, 67:444–446. doi:10.1002/jmv.10093 PMID:12116042
10. Evans AS (1976). Causation and disease: the Henle-Koch postulates revisited. *Yale J Biol Med*, 49:175–195. PMID:782050
11. Evans AS, Mueller NE (1990). Viruses and cancer. Causal associations. *Ann Epidemiol*, 1:71–92. doi:10.1016/1047-2797(90)90020-S PMID:1669491
12. Fredericks DN, Relman DA (1996). Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev*, 9:18–33. PMID: 8665474
13. Franco EL, Correa P, Santella RM *et al.* (2004). Role and limitations of epidemiology in establishing a causal association. *Semin Cancer Biol*, 14:413–426. doi:10.1016/j.semcancer.2004.06.004 PMID:15489134
14. Franco EL. Measurement errors in epidemiological studies of human papillomavirus and cervical cancer. In: Muñoz N, Bosch FX, Shah KV, Meheus A, editors. *The epidemiology of human papillomavirus and cervical cancer*. Oxford: Oxford University Press; 1992. p. 181–197.
15. Franco EL (2000). Statistical issues in human papillomavirus testing and screening. *Clin Lab Med*, 20:345–367. PMID:10863644
16. Franco EL, Rohan TE, Villa LL (1999). Epidemiologic evidence and human papillomavirus infection as a necessary cause of cervical cancer. *J Natl Cancer Inst*, 91:506–511. doi:10.1093/jnci/91.6.506 PMID: 10088620

