Antimicrobial Drug Residues in Milk and Meat: Causes, Concerns, Prevalence, Regulations, Tests, and Test Performance

J. M. MITCHELL,1* M. W. GRIFFITHS,2 S. A. MCEWEN,2 W. B. MCNAB,3 AND A. J. YEE1

1University of Guelph, Laboratory Services Division, 95 Stone Rd. West, Guelph, Ontario, Canada N1H 8J7; 2University of Guelph, Guelph, Ontario; and 3Ontario Ministry of Agriculture, Food and Rural Affairs, Education, Research and Laboratories Division, Guelph, Ontario, Canada

MS 97-87: Received 5 May 1997/Accepted 6 November 1997

ABSTRACT

This paper presents a historical review of antimicrobial use in food animals, the causes of residues in meat and milk, the types of residues found, their regulation in Canada, tests used for their detection, and test performance parameters, with an emphasis on immunoassay techniques. The development of residue detection methods began shortly after the introduction of antimicrobials to food animal production in the late 1940s. From initial technical concerns expressed by the dairy industry to the present public health and international trade implications, there has been an ongoing need for reliable, sensitive, and economical methods for the detection of antimicrobial residues in food animal products such as milk and meat. Initially there were microbial growth inhibition tests, followed by more sensitive and specific methods based on receptor binding, immunochemical, and chromatographic principles. An understanding of basic test performance parameters and their implications is essential when choosing an analytical strategy for residue testing. While each test format has its own attributes, no one test will meet all the required analytical needs. Therefore the use of a tiered or integrated system employing assays designated for screening and confirmation is necessary to ensure that foods containing violative residues are not introduced into the food chain.

Antimicrobials were first used in veterinary medicine for the treatment of mastitis in dairy cows shortly after they were developed (13, 38, 43, 45, 54, 140, 166, 170). In addition to their therapeutic value, the discovery in 1950 of the ability of antibiotics to enhance growth and feed efficiency of food animals led to their widespread use as feed supplements (75, 80, 93). Since then the use of antimicrobials has increased to such an extent that many feel that today’s intensive agricultural production techniques would be impossible without them (11, 109). The range of veterinary medicines used in or on food animals is extremely wide, ranging from teat dips to hormones. Approximately 42% of all veterinary pharmaceuticals used worldwide are used as feed additives, 19% are used as anti-infectives (e.g., antibacterials, antifungals and antivirals), 13% as parasiticides, 11% are used as biologicals and 15% represent other pharmaceuticals (107). Antimicrobials represent the largest proportion of pharmaceutical sales both in volume and dollar value of any drugs used in animal production (107).

The most commonly used antimicrobials in food animals can be grouped into five major classes. These include the beta-lactams (beta-lactams; e.g., penicillins and cephalosporins), tetracyclines (e.g., oxytetracycline, tetracycline and chlorotetracycline), aminoglycosides (e.g., streptomycin, neomycin and gentamicin), macrolides (e.g. erythromycin) and sulfonamides (e.g., sulfamethazine [SMZ]). A survey of veterinarians in the United States revealed that antibiotics were the drugs most often prescribed or used in the treatment of lactating dairy cows, mainly for mastitis therapy (149). Penicillin G was the most frequently used, and except for oxytetracycline the five most prescribed drugs were all beta-lactams approved for use in lactating dairy cattle: penicillin G, ceftiofur sodium, cloxacillin, cephalixin and ampicillin.

Antimicrobials are administered to animals by injections (e.g., intramuscular, intravenous, subcutaneous), orally in the food and water, topically on the skin and by intramammary and intrauterine infusions. Theoretically, all of these routes may lead to residues appearing in foods of animal origin such as milk, meat and eggs. It has been reported that approximately 12% of the U.S. milk supply was adulterated with beta-lactam antibiotics prior to 1962 (66). Similar results were reported in Britain in 1963 when 11% of milk samples tested were found to contain penicillin (53). In addition, controversial studies conducted in 1988 using more sensitive detection methods, indicated that 75% of North American consumer milk contained detectable levels of tetracyclines, SMZ, and other antibiotics (12, 26, 31, 81, 92, 128, 145). A survey conducted in 1969 of 5,000 samples of tissue, urine and/or feces samples collected from swine, beef cattle, veal calves, lambs and poultry at the time of slaughter in Illinois found antibiotic residues in 27%, 9%, 17%, 21% and 20%, respectively (66). Kampelmacher et al. reported in 1962 that 12% of Danish adult cattle, 58% of the calves and 23% of the swine had antibiotic residues (76). Van Schothorst reported that 77% of Danish calves and 1%
of swine and cattle had residues in 1965 (160). In France, Pitre found antibiotic residues in 4.1% of the meat samples tested in 1963 (127). A survey of Canadian slaughter animals in 1974 found that 1.9% and 3.6% of beef kidney and urine, 0.81% and 7.7% of swine kidney and urine, 0% and 9.7% of sheep kidney and urine and 0% of chicken kidneys tested positive for antibacterial agents (155). From 1974 to 1977 the violative rate for SMZ in slaughtered swine in the United States was between 10 and 13% (147).

The results of these early studies show considerable variability due to regional animal husbandry, treatment and slaughter practices and reflect the different sampling and test methodologies used. However, current data estimate that 1% of animal products in the United States and Europe contain antibiotic residues at very low levels (129).

THE ORIGIN OF RESIDUES IN MEAT AND MILK

A survey of all violative carcasses in the United States in 1993 revealed that the drugs most frequently causing residues were penicillin (20%), streptomycin (10%), oxytetracycline (10%), SMZ (9%), tetracycline (4%), gentamicin (4%) and neomycin (3%) (124). The slaughter classes most often associated with residues were culled dairy cows, veal calves and market hogs. Injectables were responsible for 46% of the violative residues in meat followed by oral administration at 20% (feed, water and bolus) and intramammary infusions at 7%. One study based on farmer opinion reported that 92% of antibiotic contamination of milk was likely due to the use of intramammary infusions (61% lactational and 31% dry cow) for the treatment of mastitis. The slaughter classes most often associated with residues were culled dairy cows, veal calves and market hogs. Injectables were responsible for 46% of the violative residues in meat followed by oral administration at 20% (feed, water and bolus) and intramammary infusions at 7%. One study based on farmer opinion reported that 92% of antibiotic contamination of milk was likely due to the use of intramammary infusions (61% lactational and 31% dry cow) for the treatment of mastitis. The slaughter classes most often associated with residues were culled dairy cows, veal calves and market hogs. Injectables were responsible for 46% of the violative residues in meat followed by oral administration at 20% (feed, water and bolus) and intramammary infusions at 7%. One study based on farmer opinion reported that 92% of antibiotic contamination of milk was likely due to the use of intramammary infusions (61% lactational and 31% dry cow) for the treatment of mastitis. The slaughter classes most often associated with residues were culled dairy cows, veal calves and market hogs. Injectables were responsible for 46% of the violative residues in meat followed by oral administration at 20%

While several factors have contributed to the residue problem such as poor treatment records or failure to identify treated animals, most violations result from the use of a drug in some manner that is inconsistent with the labelling (124, 147). This occurs primarily through not observing label withdrawal times as well as "extra-label" use of the drug. Treatments involving any other method than what is stated on the product label (e.g., different species, increased dosage, different route of administration, different frequency of treatment) are classified as extra-label usage, and withdrawal times are difficult or impossible to determine in these situations (100, 124, 147).

CONCERNS AND PUBLIC HEALTH RISKS

The initial concerns with regard to antimicrobial residues in foods were not expressed by consumers but by dairy processors who found that contaminated milk was inhibiting the starter cultures used in the production of fermented milk products as well as influencing the results of the dye reduction tests used for milk quality at the time (69, 70, 142, 143, 169). There was some concern expressed over public health hazards (e.g., allergenic), but this received little attention (48, 82). Thus, the impetus for the control and monitoring of drug residues in foods initially came from the dairy industry itself. It was not until the mid-1960s when the public became more concerned about the possible environmental and health hazards of exposure to chemical residues that more stringent controls on the use of antimicrobial drugs in food animals and tolerance levels were established (53, 82, 105, 143, 158).

Many human and animal health concerns have been expressed over the years in regard to the overuse of antibiotics in agricultural production as well as the presence of residues in the food chain. Such concerns include the potential for allergic reactions in sensitized individuals (penicillins), toxicity such as aplasia of the bone marrow (chloramphenicol), effects on the human gut microbial populations, the emergence of resistant bacteria within animals and the transfer of antibiotic resistance genes to human pathogens (47, 53, 67, 110, 158). In addition, some compounds such as the nitrofurans have been found to be animal carcinogens and mutagens in genotoxic tests. The validity of any public health threat posed by these concerns has been debated in the scientific community for over 40 years. A few cases of minor allergic reactions (e.g., skin rashes) in previously sensitized individuals to penicillin G residues in milk and meat have been documented as well as strong evidence linking widespread agricultural use of antibiotics to an increase in antibiotic resistance among animal and human pathogens (17, 37, 47, 67, 82, 110, 122, 164). Nevertheless, the public health risks of antibiotics and their metabolites in foods are difficult to define, and the presence of violative levels of residues in foods is illegal and subject to financial penalties in many countries (129).

MAXIMUM RESIDUE LIMITS

A number of national and international organizations are involved in the development of control mechanisms for the drugs used in animal production. Individual countries may follow different guidelines. These mechanisms include control of the distribution, use, determination of safe residue levels and residue detection technologies to be employed. On the international level these organizations include the Codex Alimentarius Commission, whose guidelines are set by the Codex Committee on Residues of Veterinary Drugs in Food (CCRVDF) based on the scientific advice of the Joint WHO/FAO Expert Committee on Food Additives (JECFA). Other international groups include the European Agency for the Evaluation of Medicinal Products (EMEA), Office International des Epizooties (OIE) and Consultation Mondiale de l'Industrie de la Santé Animale (COMISA). Many countries have specialist groups involved such as the Food and Drug Administration in the United States, the Bureau of Veterinary Drugs in Canada and the Veterinary Products Committee of the Ministry of Agriculture, Fisheries and Foods in the United Kingdom (153).
Limits have been established for drug residues in foods in the form of tolerances or maximum residue limits (MRLs). It should be noted that the term tolerance is used in the United States while MRL is used in Canada and the European Union, but the two terms are synonymous [15]. The term MRL may be defined as the maximum concentration of marker residue (e.g., parent compound, metabolites, etc.) resulting from the use of a veterinary drug, expressed in parts per million (ppm) or parts per billion (ppb) on a fresh weight basis, that is legally permitted or recognized as acceptable in or on food. The MRL is based on the acceptable daily intake (ADI) for that compound. The ADI is a rough estimate of the amount of a veterinary drug, expressed on a body weight basis, that can be ingested daily over a lifetime by a human being without appreciable toxicological health risk and can be considered the safety standard for that compound [14]. The following procedures are used by the Bureau of Veterinary Drugs, Human Safety Division, Health Canada for the establishment of MRLs and withdrawal periods in Canada [63, 173].

**Calculation of the ADI of a veterinary drug.** The ADI is determined by the no observable effect level (NOEL) or the dosage level (mg/kg or ppm) at which no adverse effects are observed as established by animal bioassay toxicological studies using the most sensitive testing methods available in the most sensitive animal species (e.g., teratogenicity, carcinogenicity, mutagenicity or immunopathological effects). The ADI is obtained by dividing the NOEL by a safety factor (SF) which varies from 100 to 1000 depending on the use of the drug in question and the amount and degree of toxicity data presented by the manufacturer.

$$\text{ADI (mg/kg/day)} = \frac{\text{NOEL}}{\text{SF}}$$  \hspace{1cm} (1)

**Calculation of the maximum acceptable total residue level of a veterinary drug in edible tissues.** The maximum acceptable total residue level (TRL) is calculated from the ADI and is based on the assumption that an average consumer of 60 kg body weight will consume 500 g of muscle tissue per day. The TRLs of other edible tissues are calculated by adjusting for a lower organ meat consumption and higher milk consumption as compared to muscle tissue (Table 1).

**TABLE 1. Consumption factors (relative to muscle) used for the calculation of the TRLs of veterinary drugs in Canada (63, 173)**

<table>
<thead>
<tr>
<th>Product</th>
<th>Beef</th>
<th>Pork</th>
<th>Sheep</th>
<th>Poultry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Liver</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Kidney</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Skin</td>
<td>NA</td>
<td>4</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Fat</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Milk</td>
<td>0.333</td>
<td>NA</td>
<td>0.333</td>
<td>NA</td>
</tr>
<tr>
<td>Eggs</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
</tbody>
</table>

*NA, not applicable.

For muscle:

$$\text{TRL (mg/kg)} = \frac{\text{NOEL}}{\text{SF}} \times 60 \times 2;$$  \hspace{1cm} (2)

for other products:

$$\text{TRL (mg/kg)} = \frac{\text{NOEL}}{\text{SF}} \times 60 \times 2 \times \text{CF};$$  \hspace{1cm} (3)

where:

- NOEL = no observable effect level (mg/kg);
- SF = safety factor (100 to 1000);
- 60 = 60 kg human body weight;
- 2 = 1/0.5 kg muscle consumption;
- CF = consumption factor for organ meats, milk and eggs.

**Assessment of metabolism studies.** Data on the metabolic fate of a radiolabelled veterinary drug in the target species are used to characterize and identify residues of the drug in each of the edible tissues and to establish their depletion profiles. Information on the identity, amounts and persistence of the residues in the tissues is used in selecting the target tissue and marker residue. The tissue whose total residues take the longest time to deplete below the acceptable TRL value is selected as the target tissue. The marker residue is the drug-related substance (parent compound, metabolite, etc.) that is best suited to monitor the level of total residues of toxicological concern in the target tissue.

**Determination of MRLs.** The term MRL or tolerance, frequently used by regulatory agencies, refers to the permissible level for marker residue. MRL should not be confused with TRL. TRL refers to the safe concentration of total residues, whereas the MRL is the detectable proportion of the marker residue that corresponds to the TRL value. For example, tilmicosin (Micotil, Eli Lilly and Company, Indianapolis, IN), a macrolide antibiotic prepared from tylosin, is an injectable formulation marketed and approved in Canada since 1990 for the treatment of bovine respiratory diseases in beef cattle but not for lactating dairy cattle [65]. Health Canada’s evaluation of tilmicosin established a NOEL of 4 mg/kg from a 1-year study in dogs [42]. A safety factor of 100 was then applied to give an ADI of 0.04 mg/kg/day. From this ADI value a TRL for bovine muscle of 4.8 ppm and 9.6 ppm in bovine liver was determined (see equations [2] and [3]). The liver and parent drug were then selected as the most appropriate target tissue and marker residue, respectively, and an administrative MRL of 1.6 ppm determined by a recognized analytical procedure. MRLs designated as administrative are not published in the Food and Drugs Act but may be used for regulatory enforcement purposes in Canada. The term safe level versus tolerance level has the same meaning in the United States. Drug manufacturers must provide a practical method of analysis for the marker residue in the target tissue which is capable of reliably determining that levels of marker residue are below the MRL and which can be used by regulatory agencies to monitor residue violations and non-compliance.
Establishment of withdrawal period. Data on the depletion of the marker residues in the target tissue is used to estimate the time when the residue content of the edible tissues of 99% of the total projected population of animals does not exceed the required MRL. This time is the withdrawal period or the minimal period of time between the last recommended treatment and the time of slaughter or time of collection for use as food (e.g., milk and eggs). This time allows the veterinary drug and its residues to decrease to levels below the established MRL. MRLs have been determined through studies by the various committees and included in legislation (e.g., Food and Drugs Act and Regulations in Canada, List of Codex MRLs for Vet Drugs, Official Journal of the European Communities, Code of Federal Regulations in the United States) for animal products such as meat, eggs, organs and milk (29, 30, 40, 57) (Table 2).

Concerns have recently been expressed that residue limits, especially for antibiotics, are too often set at or around the limits of analytical determination rather than on the basis of toxicology or risk assessment (153). In addition, the involvement of many organizations in the legislation of veterinary drugs has made it very difficult to standardize control practices and harmonize tolerance levels internationally. While it may be assumed that differences in tolerances levels are due to the food safety assessment of the drug, they are more often due to differences in the use of the compound, the choice of safety factors and food consumption values or the choice of target analyte used in monitoring programs and are usually insignificant in regard to food safety (15). For these reasons it has been proposed that the ADI should be used for determination of food safety rather than the MRL (14).

**TABLE 2. Maximum residue limits (MRLs) or tolerances of approved veterinary drugs for milk** (29, 40, 57)

<table>
<thead>
<tr>
<th>Drug</th>
<th>MRL or tolerance (ppb)*</th>
<th>Canada</th>
<th>European Union</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>10</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Cefotaxime</td>
<td>1,000† (A)</td>
<td>100</td>
<td>50‡ (S)</td>
<td></td>
</tr>
<tr>
<td>Cephaprin</td>
<td>20</td>
<td>—</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Cloxacillin</td>
<td>30 (A)</td>
<td>30</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Dihydrostreptomycin</td>
<td>125</td>
<td>200</td>
<td>125 (S)</td>
<td></td>
</tr>
<tr>
<td>Erythromycin</td>
<td>50</td>
<td>40</td>
<td>50 (S)</td>
<td></td>
</tr>
<tr>
<td>Neomycin</td>
<td>250 (A)</td>
<td>500</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>Novobiocin</td>
<td>125 (A)</td>
<td>—</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Oxytetracycline</td>
<td>150 (A)</td>
<td>100</td>
<td>30 (S)</td>
<td></td>
</tr>
<tr>
<td>Penicillin G</td>
<td>6</td>
<td>4</td>
<td>5 (S)</td>
<td></td>
</tr>
<tr>
<td>Streptomycin</td>
<td>125</td>
<td>125</td>
<td>125 (S)</td>
<td></td>
</tr>
<tr>
<td>Sulfadimethoxine</td>
<td>10</td>
<td>100</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Sulfadoxine</td>
<td>10 (A)</td>
<td>100</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Sulfamethazine</td>
<td>10 (A)</td>
<td>100</td>
<td>10 (S)</td>
<td></td>
</tr>
</tbody>
</table>

* A, administrative MRL (not published in Canada’s Food and Drug Act and Regulations but may be used for legislative enforcement); S, safe level (not published in the U.S. Food, Drugs and Cosmetics Act but may be used for legislative enforcement); †, parent drug and metabolites; ‡, parent drug.

**RESIDUE DETECTION METHODS**

Currently there are six types of detection methods commonly used for the detection of antimicrobial residues in foods, including microbial growth inhibition assays, microbrial receptor assays, enzymatic colourimetric assays, receptor binding assays, chromatographic methods and immunoassays.

Residue tests are either qualitative, quantitative or semiquantitative. Qualitative assays employ a predetermined cutoff value to classify samples as positive or negative relative to a specific drug concentration. Quantitative assays require that positive controls covering a wide range of drug concentrations be tested with each sample set, thus permitting residue quantitation by extrapolation from a standard curve. Such assays require precise instrumentation to measure the test response and determine the standard curve. Semiquantitative assays are similar to quantitative assays except that test results are interpreted relative to a range of drug concentrations (e.g., negative, low positive, high positive) reflected by the range of positive controls run with test samples.

Most residue assays commercially available for field applications are qualitative or semiquantitative and are classified as screening assays. A screening assay may be defined as an assay that gives a reliable and accurate indication that the analyte of interest is not present in the sample at unsafe or violative levels (35, 121, 141). This requires that screening assays be developed with a detection level optimized below the unsafe or violative levels (MRLs) so that a violative sample will have a high probability of causing a positive test result (121). Quantitative assays require much more technical expertise; therefore their primary use has been found in laboratory confirmation applications (64, 115, 135). For a summary of the development of various residue detection assays see Figure 1.

**Microbial growth inhibition assays.** The earliest methods used for the detection of antimicrobial residues in foods were based on the detection of growth inhibition of various sensitive bacterial strains. Such methods, originally developed for use in clinical medicine, were based on microbial agar diffusion tests or the inhibition of acid production or coagulation by starter organisms. Other growth inhibition assays such as conductance measurement and ATP bioluminescence have not been well received in current analytical programs (27, 62). Most of the test development work in this area was done with milk between 1947 and 1967 (46, 49, 72, 90, 97, 116, 133, 136, 168). Since that time, research has focused primarily on improving test sensitivities for a broader range of antimicrobials and improving speed, simplicity, accuracy and standardization of procedures (55, 58, 60, 77, 79, 101, 102, 106, 120, 130, 131, 143, 144, 159, 163). It was not until the late 1960s that the majority of assays were developed for the testing of tissues, primarily by modification of the existing milk testing procedures (3, 6, 8, 16, 41, 82, 96, 119, 151, 161). In fact, it was nearly 20 years after routine milk testing began in the United States that the technology became available to allow producers to test animals on the farm before sending them to slaughter (74).
FIGURE 1. Developments in antimicrobial residue testing of meat and milk.

The basic microbial inhibition assay format involves a standard culture of a test organism, usually Bacillus stearothermophilus, Bacillus subtilis, Bacillus cereus, Micrococcus luteus, Escherichia coli, Bacillus megaterium or Streptococcus thermophilus seeded in an agar or liquid growth medium which is then inoculated with a milk, urine, tissue or tissue fluid sample and incubated for periods of up to several hours. Sample can be applied directly to the medium, in stainless steel cylinders (penicylinders) or on a filter paper seed or a filter paper disk impregnated with liquid sample. During incubation the liquid diffuses into the medium and if the sample contains inhibitory substances the growth of the indicator organism will be reduced or inhibited. Depending on the format of the test the presence of an inhibitory substance is indicated by zones of growth inhibition.

The major disadvantages of microbial inhibition assays are that they are not very specific for antibiotic identification purposes, are qualitative, have limited detection levels to many antibiotics and require several hours before results are available (2.5 to 18 h). Growth inhibition tests can usually only classify residues as belonging to the β-lactam antibiotics, antibiotics other than β-lactams or sulfa drug family by use of the enzyme penicillinase or by the addition of aminobenzoic acid to the media (143, 163). In addition, many reports have stated that growth inhibition tests are subject to the effects of many natural inhibitory substances found in foods of animal origin such as lysozyme, lactoferrin, lactoperoxidase, somatic cells, complement, defensins, long-chain fatty acids, bile and lactic acid. These compounds may give false-positive test results, particularly when mastitic milk from individual cows, kidneys from pigs and urine samples are tested (18, 23, 33, 41, 59, 85, 89, 139, 157, 159, 161). Some of these effects have apparently been overcome by the use of filter paper disks in the assays, heat treatment of samples prior to testing or the use of dialysis membranes to separate the large molecular weight (MW) proteins from the smaller MW antibiotic molecules (10, 89, 108, 139, 159, 163). The primary advantages of these tests are that they are inexpensive, are easy to perform, are adaptable to screening of large numbers of samples and have a reasonably broad antimicrobial detection spectrum.

**Microbial receptor assays.** The CHARM I and II tests are qualitative microbial receptor assays for the rapid detection of β-lactams, macrolides, aminoglycosides, tetracyclines, chloramphenicol and sulfonamides in milk and tissue (25, 26, 87). Although analogous in test principle to the radioimmunoassay (RIA), by strict definition the CHARM I and II tests cannot be classified as RIAs (26). The CHARM I test for β-lactams in milk was the first Association of Official Analytical Chemists (AOAC)-recognized rapid test for the detection of β-lactams in milk with a test time of 15 min. The CHARM tests use two types of bacterial cells (Bacillus stearothermophilus binding reagent) containing either the natural receptor sites for antibiotics on or within the cells or an antibody coating (e.g., in tetracycline and chloramphenicol test kits) and a radiolabelled (14C or 3H) antibiotic (tracer reagent). Milk or tissue supernatant is added to a freeze-dried pellet of binding reagent in a test tube and the sample is mixed and incubated. During incubation any antibiotic present in the milk or tissue will bind to its specific natural or antibody receptor site on the bacterial cell. After reaction is complete the sample is mixed and incubated. At this time any unbound antibiotic is separated from the bacterial cells by centrifugation and the sample is assayed using a liquid scintillation fluid.

**Enzymatic colourimetric assays.** The Penzyme test is a qualitative enzymatic method for the rapid detection of β-lactam antibiotics in milk (83, 146, 154). Test results are available in 20 min. The test principle is based on the detection of the inactivation of an enzyme by β-lactam antibiotics. The enzyme (D-carboxypeptidase or penicillin binding protein) is present in all bacteria and is involved in the synthesis of the bacterial cell wall. β-Lactam antibiotics will bind specifically with this enzyme and inactivate it, thus
interfering with cell wall formation. This enzyme has been freeze-dried and placed in sealed vials to which the milk sample is added. After addition of 0.2 ml (200 μl) of milk sample to the vial the sample is incubated for 5 min at 47°C. During this time any β-lactams that may be present in the milk bind to the enzyme and inactivate a portion depending on the antibiotic concentration present. A reagent tablet specific for the enzyme (containing D-alanine peptide and D-amino acid oxidase) is then added to the milk sample, and the sample is incubated at 47°C for 15 min. During incubation any remaining active enzyme will react with the reagent added. The end products of the substrate and enzyme reaction (pyruvic acid and hydrogen peroxide) are measured by use of a redox colour indicator and comparison of the final colour to a colour chart provided with the kit. An orange colour (reduced) indicates a negative test result while a yellow colour (oxidized) indicates a positive test result.

Receptor binding assays. The SNAP and Delvo-X-Press tests for β-lactam antibiotics in milk are qualitative enzyme linked receptor binding assays in which β-lactams are captured by a penicillin binding protein conjugated to an enzyme (horseradish peroxidase) (4, 110). Although the test principle is similar, these systems do not use antibodies to specifically bind the β-lactams and therefore cannot be classified as immunological methods. In the first step of these tests calibrated amounts of milk and conjugate are mixed and incubated. The enzyme conjugate will bind with β-lactams that may be present in the milk sample. This mixture is then transferred to a tube (in the Delvo-X-Press method) or a plastic unit containing sample and control spots on filter paper (the SNAP device). The tubes and test spots are coated with β-lactam antibiotic. Only free enzyme conjugate will bind to this coating. Addition of enzyme substrate (colour developer) results in the formation of a blue colour, the intensity of the colour being inversely proportional to the amount of β-lactam in the sample. The absence of β-lactams in a sample results in all of the enzyme conjugate remaining unbound and available for binding to the immobilized β-lactam, while the presence of β-lactams results in a portion of the enzyme conjugate being bound and unavailable to bind with the immobilized β-lactam. Samples are declared positive or negative on the basis of a visual comparison of the intensity of colour development between the sample and control spots on the SNAP test. A positive result occurs when the sample spot is lighter than the control spot, a negative result when the sample spot is darker or equal to the control spot. Alternatively, SNAP devices may be read on a SNAP Image Reader which uses reflectance metering to compare the colour intensity of the test spots and calculates a control/sample ratio. Samples are considered positive when the ratio is greater than 1.05. The Delvo-X-Press test uses an optical density reader and compares the optical density (OD) of a standard, set at a cutoff level of 5 ppb penicillin G, with the OD of each sample tube. Samples giving an OD reading greater than zero are considered positive, while samples reading below zero are considered negative. Total assay time for both of these tests is approximately 10 min.

Chromatographic analysis. Chromatography is commonly used for separating the components of a solution. Tswett first discovered the process of liquid chromatography (LC) in 1906 when he separated the chlorophyll pigments in green leaves by passing an ether solution of these pigments through a tube of solid powdered calcium carbonate (156). Chromatography was not used extensively until 25 years later when many new applications were developed, but these early methods tended to be very slow and inefficient, had poor resolution and quantitative ability and were difficult to automate. It was not until the development of paper chromatography in the 1940s and thin layer chromatography in the 1950s that the speed and resolution of LC was greatly improved and was used more extensively. Further improvements in the chromatographic process (e.g., instrumentation) led to the development of high-performance liquid chromatography (HPLC) in the late 1960s and allowed the potential of chromatography to be realized (95).

In drug analysis, chromatography was originally used to verify drug levels in formulations, fermentation broths or biological fluids for clinical applications (115). The initial application of chromatographic methods for the detection of drug residues in foods was very limited due to the sensitivity required and poor recovery from the more complex food matrices (135). It was not until the early 1980s that residue methods were developed, primarily for the detection of β-lactams in milk and meat (112, 113, 114).

As a rule, chromatographic methods are more expensive, are more labour intensive and require more complex instrumentation and technical skills than other analytical methods. Their use as direct screening tests has thus been limited. However, the ability of chromatographic methods to specifically identify and quantitate very low levels of chemical residues has led to their use primarily as confirmatory tests for screening test positive samples (1, 2, 9, 22, 32, 61, 64, 126). It is critical that their sensitivity equal or exceed the sensitivity of screening tests; otherwise doubt will remain as to the identity of residues detected by screening tests (115, 135). Unfortunately, the development of adequate methods has not been entirely successful, although much progress has been made in recent years and methods are available for most of the major antibiotic classes (115).

There are several types of chromatographic methods currently in use for residue analysis. These include GC (gas chromatography), TLC (thin layer chromatography), TLC/BA (thin-layer chromatography/bioautography) and HPLC (high-pressure or high-performance liquid chromatography). Due to the polar, non-volatile and heat sensitive nature of most antibiotics, HPLC is the most commonly used detection method for residue analysis (135). TLC has found some use, but this method is generally used only for screening or qualitative analysis. The Sulfa On Site (SOS) Test is a TLC system used to detect SMZ in urine, serum and animal feeds which was adopted in 1988 as the official USDA method for in-plant testing of swine for SMZ residues (147).

Immunooassays. The specificity of the immune system is demonstrated by its ability to distinguish subtle differ-
ences between antigens (Ags). Antigens are defined as any substance capable of generating a specific immune response, with production of antibodies (Abs). An Ab molecule consists of two identical heavy polypeptide chains and two identical light polypeptide chains which are held together by disulphide bonds (Figure 2). Abs possess regions (paratopes) that have a specific three-dimensional geometric shape into which a corresponding part of the Ag molecule fits (epitope). This is analogous to a key in a lock. Each Ab molecule has two identical Ag binding sites. Abs have been divided into five classes based on differences in their chemical structure and biological functions (e.g., IgG, IgM, IgA, IgE and IgD). It is the specificity of the Ab-Ag reaction which forms the basis of the immunoassay. Monitoring of the Ag-Ab reaction allows the determination of Ag with a high degree of accuracy and reproducibility.

The first in vitro use of the Ag-Ab reaction was in the clinical laboratory where “agglutination” and “immunoprecipitation” reactions were extensively used for bacterial typing and toxin identification. These tests, however, were restricted to Ag-Ab reactions where the results were directly visible (171). The development of Ag and Ab chemical labelling techniques greatly increased the potential use of the immunoassay as an analytical tool as Ag-Ab reactions which were not directly detectable could now be determined using more precise methods and smaller amounts of substances could be distinguished. Examples include the development of the radioimmunoassay (RIA) in 1959 and the enzyme-linked immunosorbent assay (ELISA or EIA) in 1971 (28, 39, 172). In addition to detecting picogram or parts per trillion (ppt) amounts of a substance the RIA and EIA could be easily applied for the screening of large numbers of samples (28, 165). Since that time there has been remarkable growth in the use of immunoassay based detection systems in food analysis (7, 20, 21, 50, 51, 68, 171). With much of the current technology perfected, immunoassays have now moved out of the laboratory and into the field. Commercial test kits are presently marketed in a variety of formats for use throughout the food production industry, allowing for the rapid on-site detection of pathogens, toxins and various chemical contaminants (36, 44, 73, 94, 123, 125, 134, 152). Some examples of commercial test kits commonly used for drug residue testing in milk and tissue include the Lactek tests for milk and Cite Sulfa Trio and EZ-Screen Quik Card for various types of matrices (5, 33, 34, 103, 104, 148).

Four stages can be identified in the development of an immunological test: preparation of Ag, production and assessment of Ab, development of a suitable assay and validation of the immunoassay in relation to test samples, both spiked and incurred (20). Two types of Ags are used. Some Ags have such a low molecular weight that they cannot stimulate an immune response when injected into an animal (haptens). Abs can be raised to haptens only if the haptens are coupled to a larger protein molecule called a carrier molecule. Haptens include drugs, pesticides and microbial toxins such as aflatoxins. Other types of Ags will induce an immune response with Ab production when injected directly into an animal. These Ags are both immunogenic and antigenic and are usually large molecules such as proteins or whole microbial cells.

Two types of Ab preparations are commonly used in immunoassays, monoclonal and polyclonal. Polyclonal Ab preparations contain a mixture of Ab molecules specific for different epitopes on the same Ag. Each Ab has a different affinity and specificity for its own particular epitope. Monoclonal Abs are Ab molecules secreted from hybrid cells formed by the fusion of tumour cells with preplasma cells (84). Each clone is capable of producing large quantities of a single highly specific Ab directed against one epitope. Polyclonal Abs usually produce more sensitive immunoassays than those employing monoclonal Abs, are fairly simple to prepare and require little technical expertise (138). Abs of the IgG class are most often used in immunoassays as they are the most stable as well as the most abundant isotype found in serum following an immune response (91, 138).

Regardless of which Abs are chosen for a specific immunoassay, they should have the following characteristics (20): 1. Producibility in sufficient quantities to be of practical use; 2. Producibility with a reproducible nature (e.g., the same Ab characteristics are repeatable at a later date); 3. Stability; 4. Appropriateness for the desired immunoassay format; 5. Sufficient robustness to function under adverse conditions such as high solvent concentrations; 6. Specificity for the Ag to be detected and little or no cross-reactivity with other molecules; 7. High affinity (e.g., strength of attraction to the Ag); 8. High avidity (e.g., bonding strength to the Ag).

**FIGURE 2.** Antibody molecule. The variable light (\(V_L\)) and variable heavy (\(V_H\)) polypeptide chain regions together form the binding site for Ag. Each Ab has a unique combination of V regions. Constant (C) regions are identical in different Abs of the same class.
If these parameters are met, the potential for a satisfactory immunological method for the analyte of interest exists. Although a variety of methods can be used to detect Ag-Ab reactions, EIA has become the most popular for chemical residue detection in food due to its extreme sensitivity, simplicity and ability to screen large numbers of samples (28, 52, 150). Regardless of the target analyte, all EIA tests involve two basic steps (56). In the first step, an Ag and Ab, one which is known and one which is provided by the sample to be tested, are allowed to react. The second step, the detection of the reaction, is carried out by adding a reagent called the enzyme conjugate. The conjugate is either an Ag or Ab chemically coupled to an enzyme such as horseradish peroxidase or alkaline phosphatase. When the conjugate reacts with the analyte, a coloured product is formed. The presence or absence of colour determines the presence or absence of the unknown Ab or Ag, and permits identification.

The most popular technique for detection of Ag in a sample is called the sandwich or non-competitive EIA in which two Abs are involved that do not compete for the same epitope on the Ag molecule (Fig. 3A). Complex molecules with a number of epitopes, such as viral or bacterial agents, are detected most sensitively using sandwich EIA methodology. Direct and indirect competitive EIAs for Ag detection have also been developed (Fig. 3B and 3C).

In the direct method, the Ab is attached to a solid phase, most often a microtitre plate. Free Ags present in the sample compete with enzyme-linked Ags for a limited number of binding sites. Indirect competitive EIAs employ a solid phase coated with Ag. Free Ag present in the sample competes with the bound Ags for binding sites on an enzyme-labelled Ab. This method can also be performed by preincubating samples with enzyme-conjugated Ab added in excess. Ab binding sites that remained free can then be determined by allowing the Ab to bind to excess Ag that is bound to the solid phase. Small molecules (haptens) such as drugs, mycotoxins and pesticides that have a limited number of epitopes can be detected solely by competitive EIA systems (56, 68). The final step is the addition of the enzyme substrate. For these methods, unlike sandwich EIA, colour development is inversely proportional to Ag concentration in the sample (Figure 4).

On the basis of the relationship between colour development and concentration (e.g., the lower the drug concentration the higher the colour intensity) it is possible to design qualitative, quantitative or semiquantitative EIAs. Figure 5 shows the sequence of events that occur when a direct competitive EIA specific for drug residues tests (A) a negative sample, (B) a positive sample and (C) a sample containing residue below the “critical cutoff value” or the point at which samples are declared either positive or negative (e.g., the MRL). As shown in Figure 5, capture Ab is immobilized onto a solid support to which sample is added. If drug residue is present in the sample, it will bind to the capture Ab. Enzyme conjugate (e.g., drug bound to an enzyme) is then added which binds to any capture Ab not saturated by drug present in the sample. Enzyme substrate is then added and the colour reaction determined. A negative sample will have all of the capture Ab sites bound by enzyme conjugate, resulting in the appearance of a dark colour. A positive sample which contains sufficient analyte to saturate all of the capture Ab sites, at or above the sensitivity level of the test, will have no colour development. A sample that contains drug at a level at or below the sensitivity or cutoff point of the test will not saturate all of the capture Ab binding sites and therefore allows some enzyme conjugate binding and reduced colour development.

FIGURE 3. EIA test formats for Ag detection. (A) Non-competitive (sandwich), (B) direct competitive, (C) indirect competitive (20). E, enzyme; ag, antigen; ab₁, first (capture) antibody; ab₂, second (detecting) antibody.
ASSESSMENT OF DIAGNOSTIC TEST PERFORMANCE CHARACTERISTICS

Assay performance characteristics are best determined by using the test on field and laboratory samples. Comparison of field and laboratory data should identify any potential problems so that they can be corrected before the test is incorporated into a testing program. Although performance characteristic criteria are common to all diagnostic tests, this discussion will focus on the EIA for drug residue detection as an example.

Dose-response curves. Most commercial on-site tests for the detection of drug residues are qualitative competitive immunoassays, and hence the user will not have access to standard dose-response curves. Figure 4 shows an “ideal” curve for a competitive immunoassay. Unfortunately, such curves are rarely obtained, even in the laboratory, due to assay interference factors such as nonspecific binding, low Ab affinity, the presence of unconjugated analyte in the enzyme conjugate, nonoptimal assay conditions (e.g., temperature and pH), sample matrix effects, reagent stability, technical errors and cross-reactivity (56, 121, 138). The shape of dose-response curves can provide important information about how a particular assay is functioning. Interference factors can cause significant changes in the shape of the curve and affect assay performance such as the determination of cutoff values. Both sample matrix effects and cross-reactivity are the most frequently encountered interference factors affecting assay performance. For example, standard dose-response curves may differ significantly due to sample matrix effects for the same analyte in buffer, plasma, urine or milk (56). Similarly, if the test Ab cross-reacts with substances other than the analyte it is designed to detect, false-positive results may occur. Cross-reactivity can occur with structural variants or metabolites of the test molecule as well as related chemical substances that may be present in the test sample. Most EIAs are either highly specific to an analyte (e.g., no relative cross-reactivity greater than 1%) or show group specificity to a range of related analytes (e.g., β-lactams, tetracyclines).

Assay sensitivity and specificity. Test performance characteristics may be defined and quantified in terms of precision and accuracy as well as ruggedness, practicability, specificity and limit of reliable measurement (167). Precision refers to the consistency, repeatability and reproducibility of test results while accuracy is reflected in the test’s ability to determine the true status of a sample (99). Unfortunately, it is possible for a very precise test to consistently produce inaccurate results. Likewise, a very accurate test may have very poor precision. When the test outcome is dichotomous (e.g., positive or negative), the test

FIGURE 4. Dose-response curves for non-competitive versus competitive immunoassays. In non-competitive assays the colour response intensifies with increasing Ag concentration whereas with competitive assays the reverse occurs (56).

![Dose-response curves](image-url)

FIGURE 5. Sequential direct competitive enzyme immunoassay test. (A) Negative sample, (B) positive sample, and (C) negative sample containing analyte below the test cutoff point (56).
accuracy can be quantified in terms of sensitivity and specificity. It is important to recognize that these terms have different meanings depending on the context in which they are used (33). Analytical or laboratory sensitivity refers to the degree or level of the assay response (e.g., ppm or ppb) while analytical specificity is the ability of the test to differentiate between analytes (e.g., penicillin and tetracycline). On the other hand, diagnostic sensitivity is defined as the proportion of reference positive samples that are test positive (e.g., contain drug residue) while diagnostic specificity refers to the proportion of reference negative samples that are test negative (e.g., contain no drug residue). In both cases, it is critical that the definitions of reference positive and negative are clarified (99). While both types of sensitivity and specificity are important, it is the diagnostic sensitivity and specificity that are most often discussed in regard to diagnostic assay performance parameters.

An ideal test will distinguish between positive and negative samples 100% of the time, but it can be difficult to achieve both excellent sensitivity and specificity in the same test. Unfortunately, a gain in one is often at the expense of the other. Figure 6 is an example of a frequency distribution curve that might be obtained during field evaluations of a qualitative competitive EIA (99). As shown in the diagram there is an area of uncertainty or overlap between "high negative" and "low positive" samples (B + C). Knowledge of the magnitude of this overlap area and at what concentrations it occurs is essential when evaluating assay performance and determining critical cutoff points. For example, if one were to move the cutoff point to the left along the X axis from 1 to 3 the area represented by "C" would decrease while B would increase in size. In other words, there would be a decrease in the number of "false-negative" test results and an increase in the number of "false-positive" test results. A false-negative result wrongly declares a sample negative, when it is actually reference positive. A test with low sensitivity will lead to more false-negative results. A false-positive result mistakenly declares a sample positive, when it is actually reference negative. Due to the extreme sensitivity of many residue tests, some kit manufacturers have contended that the term false positive is incorrect and should be replaced by the term false violative or nonactionable positive (24). How the differentiation can be made using a qualitative test alone is not clear (52). A test with low specificity will lead to more false-positive results. Conversely, if the cutoff point was shifted from 1 to 2 along the X axis, the opposite would occur with a corresponding decrease in sensitivity or more false-negative results, and an increase in specificity or fewer false-positive results.

Performance parameters and decision making. Data for calculating diagnostic sensitivity and specificity can be generated by testing a number of samples using both the assay to be evaluated and an accepted "gold standard" quantitative procedure which is biologically independent of the assay being evaluated (e.g., HPLC) (99, 137). The sometimes erroneous assumption must be made that the latter always yields the correct test results (138). The samples used in the comparison should be representative "field" samples rather than "spiked" samples (20, 138). From these trials it is possible to calculate the number of times an assay will yield an incorrect result and the analyte concentrations at which such results will most likely occur. One technique used by epidemiologists to quantitatively evaluate test methods is a standard 2 X 2 table (Figure 7). Samples positive on both tests are considered true positives (TP), samples negative on both assays are considered true negatives (TN), and differing test results are classified as either false positives or false negatives in reference to the "gold standard" procedure. From these data the assay sensitivity and specificity can be calculated. With additional information on the true prevalence the predictive value of a positive test (e.g., the fraction of positive results that are true positives) and the predictive value of a negative test (e.g., the fraction of negative results that are true negatives) can also be determined (56, 99, 137). For example, if 50% of a population is reference positive, then a test with 80% sensitivity and 90% specificity will have predictive values of 89% and 82% for positive and negative test results, respectively. This can be calculated for a sample of 20,000 as follows (see Figure 7).

Reference Method

<table>
<thead>
<tr>
<th>TEST RESULTS</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Absent</td>
<td>FP</td>
<td>TN</td>
</tr>
</tbody>
</table>

Sensitivity = \( \frac{TP}{TP+FN} \times 100 \)
Specificity = \( \frac{TN}{FP+TN} \times 100 \)

PV(+) = \( \frac{TP}{TP+FP} \times 100 \)
PV(-) = \( \frac{TN}{FN+TN} \times 100 \)

FIGURE 7. Standard 2 X 2 table used to evaluate diagnostic test kit performance parameters. The test sensitivity, specificity and positive and negative predictive values can be calculated using the indicated formulas (56).
Total tests = 20,000
Prevalence of reference positives = 50%
Reference positive = 10,000
Reference negative = 10,000
TP = 10,000 × 0.80 = 8,000
FP = 10,000 − 9,000 = 1,000
FN = 10,000 − 8,000 = 2,000
TN = 10,000 × 0.90 = 9,000
PV(+) = 8,000/(8,000 + 1,000) × 100 = 89%
PV(−) = 9,000/(9,000 + 2,000) × 100 = 82%

On the other hand, if less than 1.0% of the population is expected to be true positives, which is reasonable for drug residues in foods, the positive predictive value drops from 89% to 8% while the negative predictive value increases to 99.8%. For example, if 20,000 samples were analyzed, the calculations are as follows.

Total tests = 20,000
Prevalence of reference positives = 1%
Reference positive = 200
Reference negative = 19,800
TP = 200 × 0.80 = 160
FP = 19,800 − 17,820 = 1,980
FN = 200 − 160 = 40
TN = 19,800 × 0.90 = 17,820
PV(+) = 160/(160 + 1,980) × 100 = 7.5%
PV(−) = 17,820/(17,820 + 40) × 100 = 99.8%

This implies that there is little confidence in a positive test result, because approximately 9 out of every 10 positive tests will be false positives and will not be confirmed on further testing. However, the confidence in a negative test result is high. All screening test positive results would require verification using a confirmatory assay with a very high specificity.

Therefore an understanding of the performance characteristics of an assay, as determined by field testing, such as sensitivity, specificity and expected prevalence of positives and negatives in a population are essential when considering the application of a test to each unique testing situation.

Once the test sensitivity, specificity and predictive values are known, decisions can be made as to the suitability of the test in different situations. For example, the cost of false-negative results from drug residue screening tests is difficult to estimate. The direct cost of false-positive results to producers can be substantial, however, as positive results may trigger the disposal of the commodity and the imposition of fines or penalties. False-positive test results may also have some indirect costs such as unnecessary delays in the therapeutic treatment of animals or the early culling of diseased cattle to avoid risks associated with treatment, concerns about the accuracy of drug withholding periods and increased anxiety in regard to food safety among consumers (34, 52). In these cases, due to the uncertain public health implications of drug residues in foods, it may be more appropriate to use a test with a lower sensitivity and higher specificity or to use two assays in combination: initially screening with an assay of high sensitivity, followed by confirmatory testing with an assay of high specificity (>99%) that can quantify the drug residue present, especially when large volumes of a commodity are involved (99, 137). If the quantitative assay detects a concentration greater than the level considered safe (e.g., MRL), only then should the commodity be disposed of and penalties imposed.

**CONCLUSIONS**

As the world’s population continues to grow, animal production practices will need to become more intensive and efficient, and may be accompanied by increasing demands for drug treatment. Currently, approximately 80% of all food animals receive medication for part or most of their lives. It is expected that in the future nearly all animals produced in the United States for food will have received a chemotherapeutic agent of some type (11). Therefore, the demands for reliable, simple, sensitive, rapid and low-cost methods for residue analysis of foods will continue to grow.

An ideal test would reliably detect all antimicrobial residues of concern in the food of interest at violative levels, produce very few false-negative results to protect the consumer as well as few false-positive results to protect the producer and national (or international) trade, be fast and inexpensive. Unfortunately, no such test currently exists. Therefore, it is necessary to combine methods into an “integrated system” in which a number of different tests are applied in sequence depending on the objectives of the analysis (64). For regulatory purposes such a strategy should include at least two or more independent methods (141):

1. Screening with a method optimized to prevent false-negative results and with an acceptable number of false-positive results at low cost (e.g., microbial growth inhibition tests);
2. Intermediate tests to identify the residue type (e.g., EIAs, receptor binding assays, microbial receptor assays);
3. Quantitative confirmation, with an independent method optimized to prevent false positive results (e.g., chromatography). Such methods usually have low sample throughput and a high cost.

**REFERENCES**