April 12, 2007

Meeting of the Joint Pediatric Advisory Committee and Anti-Infective Drug Advisory Committee

To Discuss

Development Pathways and Clinical Trial Designs For Products Seeking Approval for Prevention of Sequela of Shiga Toxin-producing *Escherichia coli* Infection

This FDA briefing document serves as a synopsis for the discussion of issues related to the development pathway for, and design of clinical trials for, products seeking approval for the prevention of sequela caused by Shiga toxin-producing *Escherichia coli* (STEC).

Challenges in Design of Clinical Trials for Therapeutic Products Indicated for Prevention of Sequela of STEC Infection

While the clinical features of the Hemolytic uremic syndrome (HUS) have been well described, our understanding of the pathophysiology of the disease process remains incomplete and definitive treatment is not available. It is believed that the Shiga toxins produced by STEC gain access to the circulation via the gastrointestinal tract and initiate a pathophysiological cascade that leads to HUS in some individuals, most frequently young children. In the United States the serotype that has most frequently been isolated from patients with gastrointestinal disease and HUS is O157:H7. Historically, approximately 73,000 cases of infection with O157:H7 are thought to occur each year in the United States and it is estimated that 10 to 15 percent of infected children progress to HUS.[16] More recent preliminary FoodNet data, however, suggests that the estimated annual incidence of illness due to O157:H7 in the United States has declined significantly (decreased 29% (CI=12%--42%).[6] Other STEC serotypes (non-O157:H7) may also cause illnesses similar to that caused by O157:H7 and may cause a substantial portion of HUS cases worldwide.[10,19] According to FoodNet data for 2005, the incidence rate per 100,000 population for STEC non-O157:H7 infection in the United States was reported as 0.33, but the incidence may be underestimated as protocols for isolation of STEC non-O157 at some FoodNet sites may be inadequate to detect non-O157:H7 serotypes.[6]

There is clearly a need to develop prophylactic products for prevention of HUS in patients with STEC infections. For example, patients with diarrhea/bloody diarrhea and stool positive for Stx using a rapid assay may be a population in whom prophylactic therapy administered early in disease could prevent progression to HUS. However, the design of clinical trials in this area is difficult due to a number of issues that include:

- Overall relatively low incidence, the sporadic nature of outbreaks, and the variable geographic distribution of STEC infection
- Progression of STEC infection to HUS of estimated 10-15% for O157:H7 serotype and possibly much lower for non-O157:H7 serotypes
- An incomplete understanding of the pathophysiology of the disease process
- Difficulty in identification of the appropriate therapeutic window for intervention for a given product and potential that timing from patient presentation and STEC diagnosis to the end of the therapeutic window may be relatively short (for example, therapeutics directed at circulating Stx1 and/or Stx2)

Experimental Animal Models for the Evaluation of Therapeutic Products Indicated for Shiga Toxin-Producing Infections

The ideal experimental animal model for a given clinical condition should exhibit clinical signs, pathogenesis and pathological lesions consistent with those seen in human patients with the same condition. It should be consistently repeatable in different laboratories and should be suitable for evaluation of multiple treatment modalities (e.g., antimicrobials, other drug or biologic therapies). The model should adequately address the maturity and physiological status of the clinical population.

In the case of disease caused by bacteria producing Shiga or Shiga-like toxins (Stx), the characteristics of interest include: hemorrhagic colitis, aspects of HUS including thrombocytopenia, hemolytic anemia, glomerular thrombotic microangiopathy, and sometimes central nervous system sequelae. A number of animal models have been characterized and utilized for study of these aspects of Shiga toxin-associated disease:

- The mouse has been shown to develop intestinal lesions and symptoms associated with these bacteria and associated mortality. Renal lesions are seen in the mouse, but are located exclusively in the renal tubules and not in the glomerulus as in the human.[14,20]

- The rabbit develops central nervous system lesions in areas coinciding with localization of globotriaosylceramide (Gb3) receptors, the putative target receptor for these toxins. Rabbits also develop renal lesions. Unfortunately, the rabbit model has limited utility as a model of infectious disease due to this species' particular sensitivity to alterations in normal gut flora by antimicrobials.[9]

- The baboon has also been shown to develop hemorrhagic colitis and renal lesions consistent with human HUS in response to Stx, but has practical limitations.[18]

- The piglet develops diarrheal disease and renal lesions characteristic of HUS following exposure to STEC and has been shown to develop neurological effects, as well. This species has been used successfully in multiple laboratories and has been shown to exhibit renal Gb3 receptor and lesion distribution that mimics that seen in human patients with HUS.[15]

In general, the piglet model appears to be the best of the published models for evaluation of therapeutic agents for STEC infections. It is most representative of the age and physiological status of the clinical population, it can develop hemorrhagic enteritis, HUS, and neurological effects characteristic of human disease, and has been used successfully in a number of independent laboratories.

We acknowledge the difficulty of performing clinical studies in this area; however, we believe that it is possible to develop a program that will provide data from clinical studies in humans that evaluate the safety and effectiveness of the product. FDA regulations require that Sponsors provide substantial evidence of safety and effectiveness from "adequate and well controlled studies" to support product approval/licensure. Traditionally "adequate and well controlled studies" has been interpreted to be replicative evidence of efficacy (and safety) from at least two well designed and statistically adequate clinical trials. Although alternative pathways for approval/licensure may exist (e.g., accelerated approval, animal efficacy rule) they may not be appropriate pathways, at this time, for products pursuing this indication.

At the April 12, 2007 joint Anti-Infective Drug Advisory Committee and Pediatric Advisory Committee meeting, invited speakers and FDA speakers will provide the committees summaries of:

- Regulatory approaches and considerations that necessarily impact the development plans and clinical trial designs for products seeking indications for treatment of STEC infection(s)
- The epidemiology of STEC infections
- Our current understanding of the pathophysiology and clinical course of STEC infections
- A composite STEC disease scale that has been proposed as a tool to quantify disease severity and progression.
- Animal models of STEC infection
- The feasibility of conducting statistically adequate clinical trials based on what is known about the current epidemiology of STEC infections and implications for the size of the safety database that may be available to determine risk versus benefit for a given product

In addition, the committees will hear perspectives from two Sponsors that are pursuing development of products (monoclonal antibodies targeting Stx1 and/or Stx2) for treatment of STEC infection prior to the onset of HUS.

The issues for discussion before the Advisory Committees will include:

1. The Agency does not believe that product approval for this indication should rely solely on efficacy data from animal models for approval (i.e., Animal Efficacy Rule); however, we would like the committees to consider the role animal data may provide as supportive evidence of efficacy. Is there an animal model of disease that the committee believes adequately replicates human STEC disease (e.g., pathogenesis, clinical signs, and pathological lesions) such that it that may be used to provide supportive evidence of safety and efficacy to support product approval/licensure (i.e., provide supportive data in place of one pivotal clinical safety and efficacy study)? If so, which model?

2. At this time it is anticipated that any product seeking approval/licensure for treatment of STEC infection would be studied in a clinical study(ies) of superiority design, in which the product + standard of care would be compared to standard of care alone. For products seeking to intervene in the disease process prior to the onset of HUS, what primary endpoint should used to determine efficacy?
    - Prevention of HUS only?
    - Are there alternative clinical endpoints that the committees consider clinically meaningful that may be included in a composite endpoint with prevention of HUS?

Selected References

1. Ake, et al. Early recognition of children infected with *E. coli* O157:H7, and associated volume expansion, associated with nephroprotection. Pediatrics 2005 Jun;115(6):e673-80.
2. Banatvala N, Griffin PM, Greene KD, Barrett TJ, Bibb WF, Green JH, Wells JG, and the Hemolytic Uremic Syndrome Study Collaborators. The United States National Prospective Hemolytic Uremic Syndrome Study: Microbiologic, Serologic, Clinical, and Epidemiologic Findings. J Infect Dis 2001;183:1063-1070.
3. Bitzan M and Rivière M. Clinical Proxy Markers for Shiga Toxin Diseases. Presented at the ASM Biodefence Meeting, Washington. (2006)
4. Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, Strockbine NA. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. J Infect Dis. 2005 Oct 15;192(8):1422-9.
5. Centers for Disease Control and Prevention (CDC). Importance of Culture confirmation of Shiga Toxin-producing *Escherichia coli* Infection as Illustrated by Outbreaks of Gastroenteritis – New York and North Carolina, 2005. MMWR 2006 Sept 29;55(38):1042-1045.
6. Centers for Disease Control and Prevention (CDC). Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 States, United States, 2005. MMWR 2006 Apr 14;55(14):392-5.
7. Chandler WL, Jelacic S, Boster DR, Ciol MA, Williams GD, Watkins SL, Igarashi T, Tarr PI. Prothrombotic Coagulation Abnormalities Preceding the Hemolytic-Uremic Syndrome. N Engl J Med. 2002 Jan 3;346(1):23-32.
8. Cornick NA, Jelacic S, Ciol MA, Tarr PI. *Escherichia coli* O157:H7 Infections: Discordance between Filterable Fecal Shiga Toxin and Disease Outcome. J Infect Dis 2002 Jul 1;186(1):57-63. Epub 2002 Jun 10.
9. Garcia A, Bosques CJ, Wishnok JS, Feng Y, Karalius BJ, Butterton JR, Schauer DB, Rogers AB, Fox JG. Renal Injury is a Consistent Finding in Dutch Belted Rabbits Experimentally Infected with Enterohemorrhagic *Escherichia coli*. J Infect Dis 2006;193:1125-1134.
10. Johnson KE, Thorpe CM, Sears CL. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin Infect Dis. 2006 Dec 15;43(12):1587-95. Epub 2006 Nov 9.
11. Jovanovic BD, Levy PS. A Look at the Rule of Three. The American Statistician 1997;51(2):137-139.
12. Karch H, Tarr PI, Bielaszewska M. Enterohaemorrhagic *Escherichia coli* in human medicine. Int J Med Microbiol. 2005 Oct;295:405-18.
13. Mølbak K, Mead PS, Griffin PM. Antimicrobial Therapy in Patients With *Escherichia coli* O157:H7 Infection. JAMA 2002;288(8):1014-1016.
14. Nakagawa I, Nakata M, Yamamura T, Wakisaka S, Kawabata S, Hamada S. Infection and Pathogenesis of A Murine Strain of *Escherichia coli* With Genetically Introduced Shiga Toxin Type I Operon In Conventional Mice. Microbial Pathogenesis 2002;33:63-72.

15. Pohlenz JF, Winter KR, Dean-Nystrom EA. Shiga-Toxigenic *Escherichia coli*-Inoculated Neonatal Piglets Develop Kidney Lesions That Are Comparable to Those in Humans with Hemolytic-Uremic Syndrome. Infection and Immunity 2005;73(1):612-616.
16. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. Epidemiology of *Escherichia coli* O157:H7 Outbreaks, United States, 1982-2002. Emerg Infect Dis 2005;11(4):603-609.
17. Rowe PC, Orrbine E, Lior H, Wells GA, Yetisir E, Clulow M, McLaine PN. Risk of hemolytic uremic syndrome after sporadic *Escherichia coli* O157:H7 infection: results of a Canadian collaborative study. Investigators of the Canadian Pediatric Kidney Disease Research Center. J Pediatr. 1998 May;132(5):777-82.
18. Siegler RL, Obrig TG, Pysher TJ, Tesh VL, Denkers ND, Taylor FB. Response to Shiga Toxin 1 and 2 in a Baboon Model of Hemolytic Uremic Syndrome. Pediatr Nephrol 2003;18:92-96.
19. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet. 2005 Mar 19-25;365(9464):1073-86.
20. Wadolkowski EA, Sung LM, Burris JA, Samuel JE, O'Brien A. Acute Renal Tubular Necrosis and Death o Mice Orally Infected with *Escherichia coli* Strains That Produce Shiga-Like Toxin Type II. Infection and Immunity 1990;58(12):3959-3965.
21. Prevention of Hemolytic Uremic Syndrome (HUS) Caused by Infection with Shiga Toxin-Producing Escherichia coli (STEC) with Monoclonal Antibody Therapy. Summary of Expert Panel convened by NIAID. June 19, 2002 6700B Rockledge Drive, Room 1205 Bethesda, Maryland.
22. International Conference on Harmonisation; Guidance on General Considerations for Clinical Trials [Guidance for Industry E8 General Considerations for Clinical Trials]. Federal Register. 1997 December 17;62(242).
23. Guidance for Industry E9 Statistical Principles for Clinical Trials. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) September 1998 ICH.
24. Guidance for Industry E 10 Choice of Control Group and Related Issues in Clinical Trials. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). May 2001 ICH.

Poster # 256

Thursday Feb 16, 2006

Clinical Proxy Markers for Shiga Toxin Diseases

Martin M. Bitzan[1], Marc M. Rivière[2]

[1]McGill University and Montreal Children's Hospital, Montreal, Quebec, CANADA.

[2]Caprion Pharmaceuticals, Ville St. Laurent, Montreal, Quebec CANADA.

Abstract

Background: Hemorrhagic colitis and hemolytic uremic syndrome (HUS) are serious complications of infections by Shiga toxin-producing E. coli (STEC) and S. dysenteriae type 1 resulting in significant acute and long-term human morbidity and economic losses. Up to 30% of cases require hospitalization. Because of high infectivity rates (<100 colony forming units) and environmental stability, these agents pose a potentially significant biodefense risk. Importantly, no preventive or causal treatment is presently available. Testing of novel therapies is hampered by varied incidence estimates for HUS. Because measurement of Stx in body fluids is currently not feasible, proxy markers of Shiga toxinemia are needed.

Methods: To generate a quantitative disease severity scale reflecting Shiga toxin-inflicted tissue injury, patients with STEC isolates were identified from hospital microbiology laboratory records and available medical charts screened for clinical information using a standardized questionnaire.

Results: A composite disease scale was designed to describe and measure Shiga toxin disease severity and progression. The scale was tested using our STEC disease database comprising 146 consecutive patients aged 1 - 16 y with bloody (85 %) or non-bloody diarrhea (15%) resulting in partial (5%) or complete HUS (13%). The resultant disease scale contains 4 categories: gastrointestinal disease markers, markers of inflammation and vasculopathy, evidence of hemolysis and thrombocytopenia, and CNS and other organ involvement. Linear numerical values are assigned within each category.

Conclusions: We propose the use of a clinical disease score to quantify STPB disease severity and progression as a proxy marker of Shiga toxin load and impact. The instrument is currently being validated in a prospective study in different clinical and geographic settings. It is expected to be a useful tool to analyze the severity of outbreaks of STPB infections and to measure clinical efficacy of novel therapeutics.

Introduction

STEC infections, and its major complication, the hemolytic uremic syndrome (HUS), are rare, acute diseases with significant morbidity and potentially fatal outcome. Shiga-toxin (Stx) is thought to be responsible both for STEC-associated hemorrhagic colitis and extra-intestinal tissue injury.

While the clinical diagnosis of HUS appears straight forward, there are no defined criteria to describe and grade the severity of HUS or of the preceding gastrointestinal disease. Stx blood or tissue levels (toxin "load") cannot be measured. Indirect (clinical) markers could substitute for the measurement of the toxin load, but objective definition and quantification are required.

The published literature identifies a limited set of signs and symptoms that are clearly linked to Shiga toxin-inflicted tissue injury; such events precede the appearance of HUS and include bloody diarrhea and hemorrhagic colitis, thrombocytopenia, hemolytic anemia, acute renal function degradation and/or failure and finally HUS. No specific scale is available to evaluate the severity and evolution of these Shiga-toxin mediated events (STME).

Other signs and symptoms can develop, mostly as part of HUS, that may be causally linked to Shiga-toxin inflicted tissue injury; but no clear consensus is found in the literature. These signs and symptoms include vomiting and dehydration, fever, leukocytosis, acute phase proteins, oedema and CNS symptoms, among others.

Epidemiolgy

Infections involving Shiga toxin-producing bacteria (STPB), primarily Shigella dysenteriae type 1, which plays a role as imported disease, and Shiga-toxin producing Escherichia coli (STEC), which is endemic in moderate climates. STEC and Shigella dysenteriae type 1 are included as Food Safety Threat in the list category B bioterrorism agents published by the Centers for Disease Control of the Departament of Health and Human Services of the American Government.

CDC reports an annual incidence in the USA of 110,220 cases (Mead et al. 1999). In the European Union, STPB infections are estimated to be 2.07 per 10,000 population.

Infections involving STPB occur sporadically or as outbreaks mostly in children in community settings (Tarr et al. 2005). In North and South America, the majority of cases are associated with a single serotype of E. coli (O157:H7) although multiple other serotypes can produce Shiga toxins. In Europe and Australia, the E. coli serotypes are more diverse, even though O157:H7 remains a very important component of the microbial population responsible for STEC infections.

Analyses of epidemiological risk factors or of the effect of therapeutic and preventive interventions require the measurement of defined clinical and biological parameters that would reflect toxin exposure and/or Shiga toxin-inflicted tissue injury.

Study Objectives

- To identify and test clinical and objective (laboratory) parameters that are thought to be causally related to the biological effects of Shiga toxin (STME).
- To develop a scale for the comparison of disease severity among patients with STEC infections
- To test the scale in a set of consecutive pediatric patients with STEC infection

Methods

To objectively qualify and quantify the severity of each of the Shiga-toxin mediated events and the other symptoms listed here, we extracted and adapted the relevant items of the Common Terminology Criteria Adverse Event scale (CTCAE version 3.0).

A composite disease scale was designed to describe and measure Shiga toxin disease severity and progression by grouping the different elements in 4 categories: gastrointestinal disease markers, markers of inflammation and vasculopathy, evidence of hemolysis and thrombocytopenia, and CNS and other organ involvement. Linear numerical values are assigned within each category.

The scale was then tested using our STEC disease database comprising 146 consecutive patients aged 1 - 16 y with bloody (85 %) or non-bloody diarrhea (15%) resulting in partial (5%) or complete HUS (13%). Retrospective patients chart analysis (RPCA) was done following a standard questionnaire. Data comparing scores of patients who experienced HUS versus those who did not is presented here.

Results

The results are presented by composite score of the 4 categories of the scale. Disease onset is defined as the day of first diarrheal symptom; most patients visited the pediatric ER for the first time during the third day after onset (data not shown).

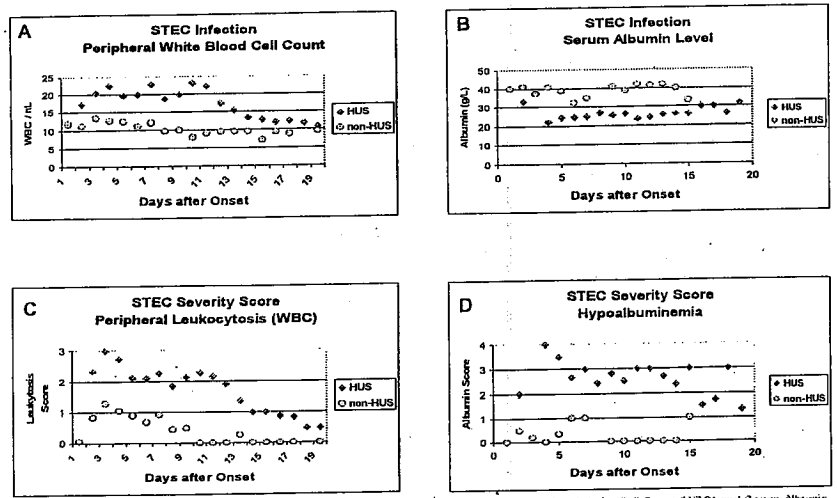


Figure 1. Validation of STEC disease markers. Shown are the mean absolute values of the peripheral White Cell Count (WBC) and Serum Albumin levels (A, B) of consecutive patients with STEC infection with relevant data (HUS = 14, Diarrhea = 79) obtained at the indicated days after onset of symptoms. Scores for degree of peripheral leukocytosis and hypoalbuminemia of the above data are presented in the lower panels (C, D). Note the separation of data points between milder (non-HUS) and more severe STEC disease (HUS) that becomes evident early after onset (days 3-5) and their merging during disease resolution (after day 14).

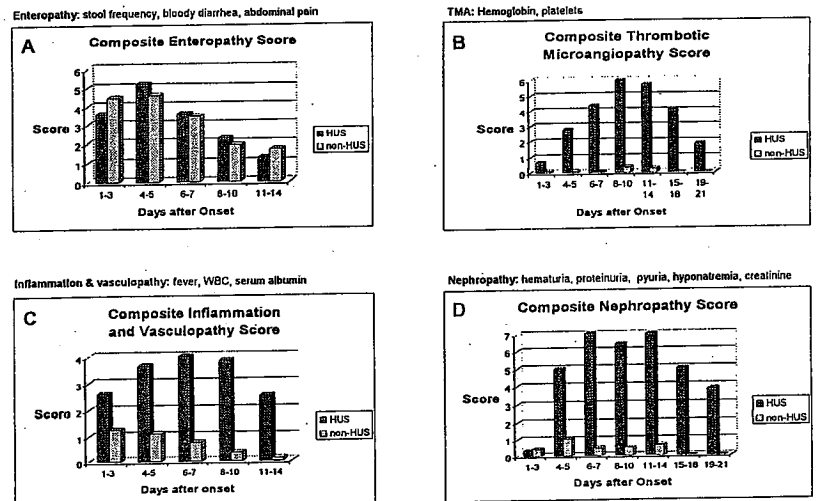


Figure 2. Composite STEC Disease Severity Scores. Clinical and laboratory data from 93 consecutive patients with bacterial culture-proven STEC O157 infection of whom relevant data were available, were scored using the STEC Disease Severity scale; for each of the chosen categories and presented correspondingly in panels A through D. Scores for patients with HUS (n = 14, including partial HUS) and with diarrhea only (n = 79) clearly separate in categories B, C and D, but are similar in category A. Note that the scale registers milder disease activity in patients with uncomplicated STEC infection (a score of zero is "normal").

Conclusion

- This work presents a novel clinical disease score for the quantitative analysis of disease acuity and severity of infections by Shiga toxin producing bacteria, specifically STEC.
- This instrument, tentatively termed STEC (or Shiga-toxin) Disease Severity Scale, addresses four distinct facets of STEC infections: colitis (enteropathy), inflammatory changes and vasculopathy, hematologic and renal changes (nephropathy) that are evident in patients with STEC colitis and HUS.
- Clear differences were observed in the degree of change in three of the four categories when patients with and without systemic complications (HUS) were compared. As expected, no quantitative differences were apparent in the enteropathy score.
- We posit that this quantitative scale is a useful instrument and propose its use in outbreaks of infections by Shiga-toxin producing bacteria as well as in epidemiological and interventional trials.

Future Work

An international prospective observational study evaluating the feasibility of data collection and relevance of the proposed scoring system and endpoints for disease follow-up has just been initiated.

Category	Grade					
	0	1	2	3	4	5
Thrombotic Micro-angiopathy						
Hemoglobin [g/l]	>115 (baseline, no schizocytosis)	>115 (drop from baseline Hb and/or schizocytosis or elevated LDH)	115 - 90 (with schizocytosis and/or elevated LDH)	89 - 65	<65 or red blood cell transfusion	-
Platelets [N/nl]	>150 (without drop from baseline)	150 – 125 (or >75 drop from baseline)	<125 – 75	<75 - 25	<25 or platelet transfusion or bleeding	-
Nephropathy						
Hematuria (dipstick RBC per HPF)	None (<3)	Small (3-5)	Moderate (6-30)	Large (>30)	-	-
Proteinuria ([g/l] or protein/creatinine ratio [g/g])	<0.3 (<0.2 g/g)	0.3 (0.2 - 0.9 g/g)	1 (1.0 - 2.9 g/g)	>3.0 (>3.0 g/g)	-	-
Pyuria [WBC/HPF]	≤2	3-10	10-50	>50	-	-
Serum creatinine	Normal for age	>upper normal for age	>2x upper normal for age	>4x upper normal for age	Dialysis	-
Hyponatremia	≥135	134-131	130-127	<127	Hyponatremic seizure	-

Category	Grade					
	0	1	2	3	4	5
Neurological Signs & Symptoms						
CNS irritability	none	Mild; easily consolable	Moderate; requiring increased attention	Severe, inconsolable	-	-
Seizure (not attributable to fever, electrolyte disorder or uremia)	none	Single, brief generalized seizure < 5 min without sequelae,	Focal seizure(s) or series of generalized, uncomplicated seizures	Complicated seizure(s) (> 5 min, post-ictal changes in mental or neurostatus)	Prolonged or repetitive seizures, that are difficult to control; status epilepticus	Death
Additional Organ Involvement						
Pancreas, exocrine enzyme deficiency and pancreatitis	none	Mild enzyme elevation (lipase, amylase) 1-2 x upper normal	Increase in stool frequency, bulk, or odor, steatorrhea; significant lipase elevation (>4x upper normal)	Pancreatitis with severe abdominal pain and enzyme elevation (lipase >4 x upper normal)	Life-threatening pancreatitis/ consequences	Death due to pancreatitis
Pancreas, endocrine Deficiency (diabetes mellitus)	none	Transient hyperglycemia / glucose intolerance	Hyperglycemia; insulin injections <48h	Diabetes mellitus; insulin treatment >48h	-	-

Category	Grade					
	0	1	2	3	4	5
Enteropathy						
Diarrhea (daily stool frequency)	No diarrhea (baseline)	<5	5-9	10-14	≥ 15	-
Abdominal pain or cramps ≤7 years of age: Wong-Baker scale (faces pain scale)	No pain (0)	Hurts "little bit" (1)	Hurts "little more / even more" (2, 3)	Hurts "whole lot" (4)	Hurts "worst" (5)	-
>8 years of age: Numeric rating scale of pain	No pain (0)	Mild pain (1-3)	Moderate pain (4-6)	Severe pain (7-9)	Unbearable pain (10)	
Bloody diarrhea (hemorrhage)	No visible blood	Occasional streaks of blood	Blood mixed with stool	Frank blood (colorectal hemorrhage)	Hemorrhage requiring colonoscopy or surgery	-
Vomiting (daily episodes)	None	1	2-5	≥6	-	-
Dehydration	None	< 5% (mild)	5% (Moderate)	10% (Severe)	≥15% (Life threatening)	-

Category	Grade					
	0	1	2	3	4	5
Inflammation & Vasculopathy						
WBC [N/nl] < 2 yrs	<15	-	15 – 16.9	17 – 21.9	>22	-
2 – 5.9 yrs	<12	12 - 13.9	14 - 16.9	17 - 21.9	>22	-
≥ 6 yrs	< 10.0	10-13.9	14 - 16.9	17 - 21.9	>22	-
CRP [mg/l]	<12	12 -19.9	20 - 27.9	28 - 35.9	>36	-
Fever	<37.5°C	37.5 - 37.9°C	38-38.9°C	>39°C	-	-
Vasculitis/ purpura (skin, mucosa)	none	Few petechiae	Peri-anal rash/erythema, localized petechiae	Generalized petechiae, purpura or bruises	-	-
Acute vascular leak syndrome and edema	none	Mild peripheral edema (periorbital, lower limb)	Moderate edema (prétibial, pre-sacral, mild ascites)	Severe edema (ascites, anasarca; pleural effusion)	Severe pulmonary edema or cardio-vascular compromise	Death due to pulmonary edema or circulatory collapse
Cerebrovascular ischemia	none	-	Migraine-type headache	Transient ischemic event or attack (TIA) ≤24 hrs duration	Cerebrovascular accident (stroke), neurological deficit or blindness	Death
Intracranial Hemorrhage, CNS	none	-	-	Mild intracranial hemorrhage (radiographic findings only)	Intracranial hemorrhage with residual neurological damage)	Death due to CNS catastrophe

three posed no substantial public health risk because they probably no longer had viremia upon arrival in the United States; although the fourth patient was likely viremic upon arrival in Minnesota in mid-May, transmission to competent local mosquito vectors in that climate was unlikely.

In early illness, the clinical features of CHIK fever can be similar to those of dengue and malaria, especially in patients without joint symptoms. In both dengue and CHIK fever, rash usually is generalized and maculopapular, but petechial rashes occur in certain dengue cases. During 1991–2004, nine confirmed or probable cases of CHIK fever were diagnosed serologically at CDC among travelers to the United States (CDC, unpublished data, 2006). Additional imported but unrecognized cases likely occurred. Clinicians should be aware of possible CHIKV infection in travelers returning from CHIK-fever–endemic or outbreak areas, particularly if an acute febrile illness with arthralgias or arthritis occurs. Suspected cases should be reported promptly to local and state public health officials and to CDC. Mosquito exposure should be strictly avoided (e.g., by staying within a screened environment and using barrier clothing and repellents) during the first week of illness to prevent infection of local mosquitoes.

In the United States, diagnostic tests for CHIKV infection are not available commercially but are available at CDC by special arrangement through state health departments. Laboratory diagnosis depends on antibody-capture IgM ELISA and plaque-reduction neutralization tests of serum. Comparative serologic tests for closely related alphaviruses (e.g., o'nyong-nyong and Sindbis viruses) should be conducted as geographically appropriate, and tests for dengue usually are indicated. Virus isolation attempts and PCR assays are performed selectively. Serologic tests should be performed on both acute- and convalescent-phase serum specimens collected at least 2 weeks apart, but clinicians should not delay submission of acute-phase samples pending collection of convalescent-phase samples. To arrange submission of specimens to CDC for diagnostic testing, clinicians should consult their state public health laboratory and CDC's Arboviral Diseases Branch (telephone, 970-221-6400). Specimen shipping and handling instructions are available at http://www.cdc.gov/ncidod/dvbid/misc/specimen-submission.htm.

References

1. Jupp PG, McIntosh BM. Chikungunya virus disease. In: Monath TP, ed. The arboviruses: epidemiology and ecology (vol 2). Boca Raton, Florida: CRC Press; 1988:137–57.
2. Depoortere E, Coulombier D, European Centre for Disease Prevention and Control Chikungunya Risk Assessment Group. Chikungunya risk assessment for Europe: recommendations for action. Euro Surveill 2006;11:E060511.2. Available at http://www.eurosurveillance.org/ew/2006/060511.asp#2.
3. Parola P, de Lamballerie X, Jourdan J, et al. Novel chikungunya virus variant in travelers returning from Indian Ocean islands. Emerg Infect Dis 2006;12:1493–9. Available at http://www.cdc.gov/ncidod/EID/vol12no10/pdfs/06-0610.pdf.
4. Ravi V. Re-emergence of chikungunya virus in India. Indian J Med Microbiol 2006;24:83–4.
5. Ligon BL. Reemergence of an unusual disease: the chikungunya epidemic. Semin Pediatr Infect Dis 2006;17:99–104.
6. Public Health Agency of Canada. Notices and international reports. Outbreak of chikungunya virus: south west Indian Ocean and India. Available at http://www.phac-aspc.gc.ca/tmp-pmv/2006/chiku060526_e.html.
7. Caribbean Epidemiology Centre (CAREC). Chikungunya in Martinique and French Guyana. CAREC surveillance report. Port of Spain, Trinidad & Tobago: Pan American Health Organization; 2006;26:17. Available at http://www.carec.org/pdf/csrjune2006.pdf.
8. Retuya TJA Jr, Ting DL, Dacula BD, et al. Chikungunya fever outbreak in an agricultural village in Indang, Cavite, Philippines. Philippine Journal of Microbiology and Infectious Diseases 1998;27:93–6.
9. Laras K, Sukri NC, Larasati RP, et al. Tracking the re-emergence of epidemic chikungunya virus in Indonesia. Trans R Soc Trop Med Hyg 2005;99:128–41.
10. Robillard PY, Boumahni B, Gerardin P, et al. Vertical maternal fetal transmission of the chikungunya virus. Presse Med 2006;35:785–8.

Importance of Culture Confirmation of Shiga Toxin-producing *Escherichia coli* Infection as Illustrated by Outbreaks of Gastroenteritis — New York and North Carolina, 2005

Escherichia coli O157:H7 and other strains of *E. coli* that produce Shiga toxin are collectively known as Shiga toxin-producing *E. coli* (STEC). The current outbreak of STEC O157 infections associated with eating fresh spinach illustrates the importance of obtaining isolates to identify the source of the infections (*1*). Laboratory methods that do not require bacterial culture of stool specimens to identify STEC are being used increasingly by clinical diagnostic laboratories, sometimes without subsequent confirmation of a strain by isolating it in culture. This report describes findings from outbreaks of gastroenteritis in 2005 in New York and North Carolina in which clinical diagnostic laboratories initially used only non-culture methods to detect Shiga toxin (Stx). The findings highlight the importance of confirmation of Stx-positive stool specimens by bacterial culture for timely and reliable identification of STEC infections, including *E. coli* O157 and non-O157 STEC, to enable implementation of appropriate public health actions. An important part of that identification is determining the serotype of all STEC isolates and the subtype of STEC O157 strains so that outbreaks can be detected and traced back to sources.

New York

During August 28–September 13, 2005, a total of 52 (2.4%) of 2,160 inmates at a state correctional facility reported diarrhea, including 17 (33%) with bloody diarrhea. Nineteen inmates were treated at the prison infirmary; three were hospitalized for an average of 1.8 days. Stool specimens from these three inmates tested positive for Stx by enzyme immunoassay (EIA) at a clinical diagnostic laboratory. Subsequently, stool specimens collected from 21 ill inmates were submitted to the New York State Department of Health (NYSDOH)-Wadsworth Center. Stool specimens were inoculated to *E. coli* enrichment broth and sorbitol MacConkey agar (SMAC), a selective medium used to screen for STEC serotype O157:H7 because this serotype, unlike most *E. coli* (and unlike most STEC), does not ferment sorbitol. Sixteen of the stool enrichment broths, when tested by polymerase chain reaction (PCR), were positive for the Shiga toxin 1 gene (*stx1*) but negative for STEC O157-specific DNA; 13 of the SMAC agar plates demonstrated growth of sorbitol-fermenting *E. coli* colonies that also were positive for *stx1* by PCR and did not agglutinate with commercial latex reagents for STEC serogroups O26, O91, O103, O111, O128, O145, and O157. Isolates from three patients were sent to CDC and determined to be STEC serotype O45:nonmotile (NM) (one patient had both STEC serotype O45:NM and O45:H2). These STEC O45 isolates were indistinguishable by pulsed-field gel electrophoresis (PFGE) using *Xba*I and *Bln*I restriction endonucleases.

The source of the outbreak likely was an ill food worker. Control measures included enhanced surveillance for additional illness and reminders of the need for exclusion of the infected food worker from food service or other jobs with increased risk for transmission until his specimens no longer tested positive for STEC.

North Carolina

On November 10, 2005, the Davidson County Health Department received a report of non-bloody diarrheal illness in an infant aged 6 months who attended a day care center. Diarrhea was reported in four additional day care center attendees and three family members of the index patient.

An enrichment broth from a stool specimen from the index patient tested positive for Stx by EIA at a clinical diagnostic laboratory. After a delay of some days, the laboratory sent the enrichment broth culture of this stool specimen to the North Carolina State Laboratory for Public Health, where neither STEC O157 nor STEC serogroups O26, O45, O103, O121, O111, or O145 were isolated. The enrichment broth was then sent to CDC, where it again tested positive for Stx by EIA, but PCR tests of the enrichment broth at CDC were negative for *stx1* and *stx2*. The enrichment broth was plated on SMAC, and PCR tests of both a sweep of growth from the plate and of 10 sorbitol-fermenting colonies were negative for *stx1* or *stx2*. Subsequently, the North Carolina State Laboratory for Public Health performed additional testing on stool specimens from five ill persons, including the index case; each tested positive for norovirus by reverse transcription-PCR.

In response to the initial Stx-positive report, public health control measures appropriate for STEC had been instituted, including exclusion of the index case from the day care center pending receipt of two STEC-negative cultures of stool specimens collected at least 24 hours apart. These exclusion measures also had been enforced for the four other ill children in the day care center. When the outbreak was determined to have been caused by norovirus, not STEC, control measures were revised, and the ill children were allowed to return to the day care center after they became asymptomatic.

Reported by: *R Atkinson, PhD, G Johnson, T Root, T Halse, D Wroblewski, MS, New York State Dept of Health. M Davies, MD, A Byrd, L Long, MSA, Davidson County Health Dept, North Carolina. L Demma, PhD, F Angulo, DVM, PhD, C Bopp, MS, P Gerner-Smidt, MD, PhD, N Strockbine, PhD, K Greene, B Swaminathan, PhD, P Griffin, MD, Div of Foodborne, Bacterial, and Mycotic Diseases, National Center for Zoonotic, Vector-Borne, and Enteric Diseases (proposed); J Schaffzin, MD, PhD, B Goode, MPH, EIS officers, CDC.*

Editorial Note: The New York outbreak demonstrates that Stx testing of stool specimens from patients with diarrhea by clinical diagnostic laboratories can facilitate detection of outbreaks of non-O157 STEC. However, as the North Carolina outbreak demonstrates, occasional false-positive results from the Stx EIA test can result in inappropriate and unnecessary public health action. These two outbreaks illustrate the importance of culture confirmation of Stx EIA-positive specimens.

Although STEC O45 is an important cause of sporadic non-O157 STEC infections in the United States, the cases in New York represent the first outbreak of STEC O45 infections ever identified in the United States (*2*). During 1983–2002, public health laboratories submitted 940 non-O157 STEC isolates to CDC, of which 7% were identified as O45, making it the fifth most commonly isolated non-O157 STEC serogroup during that period (*2*).

The outbreak in North Carolina illustrates the importance both of rapid culturing of all Stx-positive broths specifically for STEC O157 and of rapid culture confirmation of Stx-positive specimens. The initial Stx-positive result prompted public health actions that, in retrospect, placed an unnecessary burden on patients, day care center staff, and public health officials. Had a culture for *E. coli* O157 been performed simultaneously with the EIA or after the EIA was determined

positive, the negative result might have prompted investigation for norovirus sooner. Once the EIA result was determined falsely positive, and the true etiology of the outbreak was determined, control measures appropriate for norovirus were instituted. This is not the first time that an outbreak of norovirus infections was mistakenly attributed to STEC (*3*).

In 2000, non-O157 STEC infections became nationally notifiable. However, few non-O157 STEC infections are detected because most clinical diagnostic laboratories do not test stool specimens routinely for these organisms (*4,5*). No selective agar medium exists for isolation of non-O157 STEC. SMAC and other sorbitol-containing selective media such as cefixime-tellurite SMAC (CT-SMAC), Rainbow Agar O157, and CHROmagar O157 enhance isolation of STEC O157 because strains of this serotype typically do not ferment sorbitol or produce beta-D-glucuronidase. However, most non-O157 STEC strains ferment sorbitol and are phenotypically indistinguishable from other *E. coli* strains present in the normal intestinal flora. Non-O157 STEC infections can be diagnosed by use of EIA, PCR, or cell culture methods to detect free Stx or the *stx1* or *stx2* genes in stool or enrichment broths. EIA testing of broth cultures, rather than the stool specimens themselves, is recommended because the amount of free fecal Stx in stools often is low (*6*). Alternatively, production of Stx or the presence of Stx gene sequences can be demonstrated by selecting colonies from plating media and testing them by EIA or PCR. The development of commercial Stx EIA kits has allowed clinical diagnostic laboratories to easily screen stool specimens for STEC independent of serotype. If the index of clinical suspicion for STEC O157 is high, the stool specimen should be tested simultaneously by Stx EIA and by bacterial culture on a sorbitol-containing medium such as SMAC (*7*). Virulence factors strongly associated with the development of hemolytic uremic syndrome (HUS) are almost always present in STEC O157, but less frequently in non-O157 STEC (*8*). The majority of clinical diagnostic laboratories cannot determine the virulence profile of STEC but can identify an STEC O157 infection. Therefore, early diagnosis of at least STEC O157 is important to identify patients at highest risk for HUS. Treatment with parenteral-volume expansion early in the course of STEC O157 infection can decrease renal injury and improve patient outcome (*9*).

Clinical diagnostic laboratories should strongly consider including STEC O157 in their routine bacterial enteric panel (with *Salmonella, Shigella,* and *Campylobacter* spp.). If bacterial culture for STEC O157 is not performed in parallel with EIA, Stx-positive broths should be inoculated to a selective isolation medium, such as SMAC agar, and any resulting sorbitol-negative colonies should be tested with O157 antiserum or latex reagent. All confirmed and presumptive STEC O157 isolates and Stx-positive broths that do not yield STEC O157 should be forwarded to a public health laboratory as soon as possible for confirmatory testing and further genetic characterization. STEC O157 isolates should be confirmed, characterized, and tested by PFGE, and the pattern promptly entered into the PulseNet database. At the public health laboratory, the broth should be subcultured to selective agar and a representative sample of sorbitol positive and negative colonies tested by Stx EIA or PCR for *stx1* and *stx2* genes. Non-O157 STEC isolates can be tested using commercial antisera for the most common non-O157 STEC serogroups (O26, O45, O103, O111, O121, and O145) and should be sent to the CDC *E. coli* Reference Laboratory for complete serotyping and further genetic characterization, including PFGE.

To facilitate investigation of possible outbreaks, clinicians should inform health departments about clusters of patients with bloody diarrhea or HUS, and clinical diagnostic laboratories should follow recommended procedures for identification of STEC (Box). Screening stool specimens by clinical diagnostic laboratories for Stx using EIA, subsequent bacterial culture of Stx-positive specimens using SMAC, and forwarding enrichment broths from Stx-positive specimens that do not yield STEC O157 to state or local public health laboratories, are crucial steps for public health surveillance of STEC infections. With this coordinated approach, accurate laboratory data can be combined with epidemiologic information to ensure prompt diagnosis and treatment of STEC O157 infections, improved diagnostic accuracy, and improved detection of outbreaks caused by non-O157 STEC.

Acknowledgments

The findings in this report are based, in part, on contributions by M Caldwell, MD, A Evans, S Ireland, M DeFabio, Dutchess County Dept of Health, Poughkeepsie; L Klopf, MSEd, C Metzler, New York State Dept of Correctional Svcs; N Dumas, T Quinlan, L Armstrong, J Edwards; K Musser, PhD, R Limberger, PhD, New York State Dept of Health-Wadsworth Center; R Gallo, B Devine, New York State Dept of Health. B Jenkins, C Hartley, L Wolf, North Carolina State Laboratory of Public Health. S Rolando, MHS, Assoc of Public Health Laboratories, Silver Springs, Maryland. KG Holt, DVM, Food Safety and Inspection Svc, US Dept of Agriculture.

References

1. CDC. Ongoing multistate outbreak of *Escherichia coli* serotype O157:H7 infections associated with consumption of fresh spinach—United States, September 2006. MMWR 2006;55:1045–6.
2. Brooks JT, Sowers EG, Wells JG, et al. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. J Infect Dis 2005;192:1422–9.
3. CDC. University outbreak of calicivirus infection mistakenly attributed to Shiga toxin-producing *Escherichia coli* O157:H7—Virginia, 2000. MMWR 2001;50:489–91.

BOX. Recommendations for laboratory identification of Shiga toxin-producing *Escherichia coli* (STEC)

- Health-care providers should notify clinical diagnostic laboratories when STEC O157 infection is suspected (e.g., because of bloody diarrhea or hemolytic uremic syndrome) so that appropriate testing methods can be applied.
- Clinical diagnostic laboratories should strongly consider including STEC O157 in their routine bacterial enteric panel (with *Salmonella, Shigella,* and *Campylobacter*).
- The best way to identify all STEC infections is to screen all stool samples submitted for routine enteric bacterial testing for Shiga toxins (Stxs) using enzyme immunoassay (EIA) or polymerase chain reaction. Ideally, the clinical diagnostic laboratory should culture simultaneously for STEC O157 (e.g., on sorbitol MacConkey agar). Simultaneous culture facilitates rapid diagnosis and treatment of patients with STEC O157 infection and rapid subtyping by public health laboratories; such rapid action is most important when the index of clinical suspicion for STEC O157 is high.
- Clinical diagnostic laboratories that use an Stx EIA but do not perform simultaneous culture for STEC O157 should culture all Stx-positive broths for STEC O157 as soon as possible and rapidly forward these isolates to a state or local public health laboratory for confirmation and subtyping.
- When an Stx-positive broth does not yield STEC O157, the broth culture should be quickly forwarded to the state or local public health laboratory for identification of non-O157 STEC.
- State and local public health laboratories should confirm the presence of Stx in broths sent from clinical laboratories and should attempt to obtain an STEC isolate. All non-O157 STEC isolates should be sent by public health laboratories to CDC for confirmation and further characterization.

4. Klein EJ, Stapp JR, Clausen CR, et al. Shiga toxin-producing *Escherichia coli* in children with diarrhea: a prospective point-of-care study. J Pediatr 2002;141:172–7.
5. Voetsch AC, Van Gilder TJ, Angulo FJ, et al. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin Infect Dis 2004;38(Suppl 3):S127–34.
6. Cornick NA, Jelacic S, Ciol MA, et al. *Escherichia coli* O157:H7 infections: discordance between filterable fecal Shiga toxin and disease outcome. J Infect Dis 2002;186:57–63.
7. Cohen MB. Shiga toxin-producing *E. coli*: two tests are better than one. J Pediatr 2002;141:155–6.
8. Eklund M, Leino K, Siitonen A. Clinical *Escherichia coli* strains carrying stx genes: stx variants and stx-positive virulence profiles. J Clin Microbiol 2002;40:4585–93.
9. Ake JA, Jelacic S, Ciol MA, et al. Relative nephroprotection during *Escherichia coli* O157:H7 infections: association with intravenous volume expansion. Pediatrics 2005;115:e673–80.

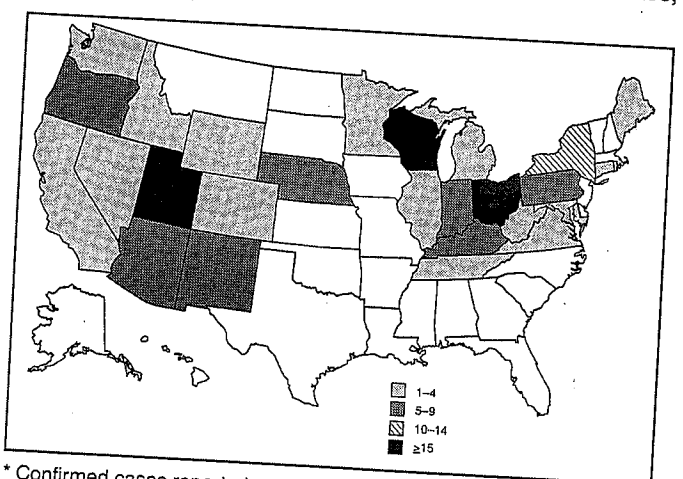
Ongoing Multistate Outbreak of *Escherichia coli* serotype O157:H7 Infections Associated with Consumption of Fresh Spinach — United States, September 2006

On September 26, this report was posted as an MMWR *Dispatch on the* MMWR *website* (http://www.cdc.gov/mmwr).

On September 13, 2006, CDC officials were alerted by epidemiologists in Wisconsin and Oregon that fresh spinach was the suspected source of small clusters of *Escherichia coli* serotype O157:H7 infections in those states. On the same day, New Mexico epidemiologists contacted Wisconsin and Oregon epidemiologists about a cluster of *E. coli* O157:H7 infections in New Mexico associated with fresh spinach consumption. Wisconsin public health officials had first reported a cluster of *E. coli* O157:H7 infections to CDC on September 8. On September 12, CDC PulseNet had confirmed that the *E. coli* O157:H7 strains from infected patients in Wisconsin had matching pulsed-field gel electrophoresis (PFGE) patterns and identified the same pattern in patient isolates from other states. This report describes the joint investigation and outbreak-control measures undertaken by state public health officials, CDC, and the Food and Drug Administration (FDA). This investigation and additional case finding are ongoing.

As of September 26, a total of 183 persons infected with the outbreak strain of *E. coli* O157:H7 had been reported to CDC from 26 states (Figure 1). Among the ill persons, 95 (52%) were hospitalized, 29 (16%) had hemolytic uremic syndrome (HUS), and one person died. The deaths of two other patients possibly related to this outbreak are under

FIGURE 1. Number of confirmed cases (N = 183)* of *Escherichia coli* serotype O157:H7 infection, by state — United States, September 2006



* Confirmed cases reported as of 1:00 p.m. EDT on September 26, 2006.

CDC Home | Search | Health Topics A-Z

MMWR Weekly

April 14, 2006 / 55(14);392-395

Preliminary FoodNet Data on the Incidence of Infection with Pathogens Transmitted Commonly Through Food --- 10 States, United States, 2005

Foodborne illnesses are a substantial health burden in the United States (*1*). The Foodborne Diseases Active Surveillance Network (FoodNet) of CDC's Emerging Infections Program collects data from 10 U.S. states* regarding diseases caused by enteric pathogens transmitted commonly through food. FoodNet quantifies and monitors the incidence of these infections by conducting active, population-based surveillance for laboratory-confirmed illness (*2*). This report describes preliminary surveillance data for 2005 and compares them with baseline data from the period 1996--1998. Incidence of infections caused by *Campylobacter*, *Listeria, Salmonella,* Shiga toxin--producing *Escherichia coli* O157 (STEC O157), *Shigella,* and *Yersinia* has declined, and *Campylobacter* and *Listeria* incidence are approaching levels targeted by national health objectives (*3*) (Table). However, most of those declines occurred before 2005, and *Vibrio* infections have increased, indicating that further measures are needed to prevent foodborne illness.

In 1996, FoodNet began active, population-based surveillance for laboratory-confirmed cases of infection from *Campylobacter*, *Listeria*, *Salmonella*, STEC O157, *Shigella*, *Vibrio*, and *Yersinia*. In 1997, FoodNet added surveillance for cases of *Cryptosporidium* and *Cyclospora* infection. In 2000, FoodNet began collecting data on STEC non-O157 and comprehensive information on hemolytic uremic syndrome (HUS). FoodNet personnel ascertain cases through contact with all clinical laboratories in their surveillance areas. HUS surveillance is conducted through a network of pediatric nephrologists and infection-control practitioners. In addition, eight states review hospital discharge data to ascertain HUS cases. Because of the time required for review of hospital records, this report contains preliminary 2004 HUS data.

During 1996--2005, the FoodNet surveillance population increased from 14.2 million persons (5% of the U.S. population) in five states to 44.5 million persons (15% of the U.S. population) in 10 states. Preliminary incidence for 2005 was calculated using the number of laboratory-confirmed infections and dividing by 2004 population estimates. Final incidence for 2005 will be reported when 2005 population estimates are available from the U.S. Census Bureau.

2005 Surveillance

In 2005, a total of 16,614 laboratory-confirmed cases of infections in FoodNet surveillance areas were identified, as follows: *Salmonella* (6,471 cases), *Campylobacter* (5,655), *Shigella* (2,078),

Cryptosporidium (1,313), STEC O157 (473), *Yersinia* (159), STEC non-O157 (146), *Listeria* (135), *Vibrio* (119), and *Cyclospora* (65). Overall incidence per 100,000 population was 14.55 for *Salmonella,* 12.72 for *Campylobacter,* 4.67 for *Shigella,* 2.95 for *Cryptosporidium,* 1.06 for STEC O157, 0.36 for *Yersinia,* 0.33 for STEC non-O157, 0.30 for *Listeria,* 0.27 for *Vibrio*, and 0.15 for *Cyclospora*. Substantial variation occurred across surveillance sites (Table). In 2004, FoodNet identified 44 cases of HUS in children aged <15 years (rate: 0.49 per 100,000 children); 30 (68%) of these cases occurred in children aged <5 years (rate: 0.94).

Of the 5,869 (91%) *Salmonella* isolates serotyped, six serotypes accounted for 61% of infections, as follows: Typhimurium, 1,139 (19%); Enteritidis, 1,080 (18%); Newport, 560 (10%); Heidelberg, 367 (6%); Javiana, 304 (5%); and a monophasic serotype identified as *Salmonella* I 4,[5],12:i:-, 154 (3%). Among 109 (92%) *Vibrio* isolates identified to species level, 59 (54%) were *V. parahaemolyticus,* and 15 (14%) were *V. vulnificus*. FoodNet also collected data on 145 STEC non-O157 isolates that were tested for O-antigen determination; 117 (81%) had an identifiable O antigen, including O26 (37 [32%]), O103 (36 [31%]), and O111 (23 [20%]); 28 isolates did not react with the typing antisera used.

In 2005, FoodNet sites reported 205 foodborne disease outbreaks to the national Electronic Foodborne Outbreak Reporting System; 121 (59%) were associated with restaurants. Etiology was reported for 159 (78%) outbreaks; the most common etiologies were norovirus (49%) and *Salmonella* (18%).

Comparison of 2005 Data with 1996--1998

A main-effects, log-linear Poisson regression model (negative binomial) was used to estimate statistically significant changes in the incidence of pathogens. This model accounts for the increase in the number of FoodNet sites and its surveillance population since 1996 and for variation in the incidence of infections among sites (*2*). The average annual incidence for 1996--1998 (1997--1998 for *Cryptosporidium*), the first 3 years of FoodNet surveillance, was used as the baseline period. For HUS surveillance, 2000--2001 was used as the baseline. The estimated change in incidence (relative rate) between the baseline period and 2005 was calculated, along with a 95% confidence interval (CI).

The estimated annual incidence of several infections declined significantly from 1996--1998 to 2005 (Figure 1). The estimated incidence of infection with *Yersinia* decreased 49% (CI = 36%--59%), *Shigella* decreased 43% (CI = 18%--60%), *Listeria* decreased 32% (CI = 16%--45%), *Campylobacter* decreased 30% (CI = 25%--35%), STEC O157 decreased 29% (CI = 12%--42%), and *Salmonella* decreased 9% (CI = 2%--15%). Although *Salmonella* incidence decreased overall, of the five most common *Salmonella* serotypes, only the incidence of *S.* Typhimurium decreased significantly (42% [CI = 34%--48%]). The estimated incidence of *S.* Enteritidis increased 25% (CI = 1%--55%), *S.* Heidelberg increased 25% (CI = 1%--54%) and *S.* Javiana increased 82% (CI = 14%--191%). The estimated incidence of *S.* Newport increased compared with the baseline, but the increase was not statistically significant (Figure 2). The estimated incidence of postdiarrheal HUS in children aged <5 years decreased 45% in 2004 compared with 2000--2001; whether this trend is significant could not be determined, partly because the limited time span does not provide enough data to evaluate a Poisson regression model. The estimated incidence of *Vibrio* increased 41% (CI = 3%--92%) compared with the baseline, whereas the estimated incidence of *Cryptosporidium* infections did not change significantly.

Reported by: *D Vugia, MD, California Dept of Health Svcs. A Cronquist, MPH, Colorado Dept of Public Health and Environment. J Hadler, MD, Connecticut Dept of Public Health. M Tobin-D'Angelo, MD, Div of Public Health, Georgia Dept of Human Resources. D Blythe, MD, Maryland*

Dept of Health and Mental Hygiene. K Smith, DVM, Minnesota Dept of Health. K Thornton, MD, Institute for Public Health, Univ of New Mexico Health Sciences Center, Albuquerque. D Morse, MD, New York State Dept of Health. P Cieslak, MD, Oregon State Public Health. T Jones, MD, Tennessee Dept of Health. K Holt, DVM, Food Safety and Inspection Svc, US Dept of Agriculture. J Guzewich, MPH, Center for Food Safety and Applied Nutrition, Food and Drug Admin. O Henao, PhD, E Scallan, PhD, F Angulo, DVM, P Griffin, MD, R Tauxe, MD, Div of Foodborne, Bacterial and Mycotic Diseases, National Center for Zoonotic, Vector-Borne and Enteric Diseases; E Barzilay, MD, EIS Officer, CDC.

Editorial Note:

In 2005, compared with the 1996--1998 baseline period, significant declines occurred in the estimated incidence of *Campylobacter*, *Listeria*, *Salmonella*, *Shigella*, STEC O157, and *Yersinia* infections. Several important food safety initiatives (*1*) might have contributed to the declines, indicating progress toward meeting the national health objectives (Table) (*3*). However, most progress occurred before 2005. Most of the decline in *Campylobacter* incidence occurred by 2001, with continued small decreases since then. The incidence of *Listeria* infections in 2005 is higher than its lowest point in 2002. Of the five most common *Salmonella* serotypes, only Typhimurium has declined, with most of the decline occurring by 2001. Most of the decline in STEC O157 incidence occurred during 2003 and 2004. The observed sustained increase in *Vibrio* incidence indicates that additional efforts are needed to prevent *Vibrio* infections. Oysters are the most important source of human *Vibrio* infections, and most human infections can be prevented by not eating raw or undercooked oysters. Measures that reduce *Vibrio* contamination of oysters also prevent illness.

Food animals are the most important source of human *Salmonella* infections. Transmission of *Salmonella* to humans can occur via various food vehicles, including eggs, meat, poultry, and produce, and via direct contact with animals and their environments. Testing by the U.S. Department of Agriculture, Food Safety and Inspection Service (FSIS) at slaughter and processing plants has demonstrated declines in *Salmonella* contamination of ground beef since 1998 (*4*). However FSIS recently announced a sustained increase in chicken-broiler carcasses testing positive for *Salmonella* during 2002--2005 and subsequently launched an initiative to reduce *Salmonella* in raw meat and poultry products (*4,5*). Although sources of infection with the most common *Salmonella* serotypes have been identified (e.g., food animals), further investigation is needed to identify sources for emerging *Salmonella* serotypes, such as Javiana and I 4,[5],12:i:-, a monophasic serotype that resembles *S.* Typhimurium except that it has no phase 2 flagellar antigen and has previously been misclassified as Group B *Salmonella* or *S.* Typhimurium (*6*).

Large outbreaks with multiple laboratory-confirmed cases can distort underlying trends in incidence. For example, the incidence of *Cryptosporidium* infections increased substantially from 2004 to 2005 because of a large outbreak associated with visits to a recreational water park in New York (P Smith, MD, New York State Department of Health, personal communication, 2006).

The findings in this report are subject to at least four limitations. First, FoodNet relies on laboratory diagnoses, but many foodborne illnesses are not diagnosed by clinical laboratories. Second, protocols for isolation of certain enteric pathogens (e.g., STEC non-O157) in clinical laboratories vary and are not uniform within and among FoodNet sites (*7*); others (e.g., norovirus) cannot readily be identified by clinical laboratories. Third, reported illnesses might have been acquired through nonfoodborne sources, and reported incidence rates do not reflect foodborne transmission exclusively. Finally, although the FoodNet surveillance population is similar to the U.S. population (*2*), the findings might not be generalizable to the entire U.S. population.

Much remains to be done to reach the national health objectives for foodborne illnesses. Enhanced measures are needed to understand and control pathogens in animals and plants, to reduce or prevent contamination during processing, and to educate consumers about risks and prevention measures. Such measures can be particularly focused when the source of human infections (i.e., animal reservoir species and transmission route) are known. The declines in the incidence of STEC O157 infections observed in recent years suggest that coordinated efforts by regulators and industry have been effective in reducing contamination and illness related to ground beef (*8,9*).

Consumers can reduce their risk for foodborne illness by following safe food-handling recommendations and by avoiding consumption of unpasteurized milk and unpasteurized milk products, raw or undercooked oysters, raw or undercooked eggs, raw or undercooked ground beef, and undercooked poultry (additional information on food safety for consumers is available at http://www.fightbac.org). Other effective prevention measures, such as pasteurization of in-shell eggs, irradiation of ground meat, and pressure treatment of oysters, can also decrease the risk for foodborne illness.

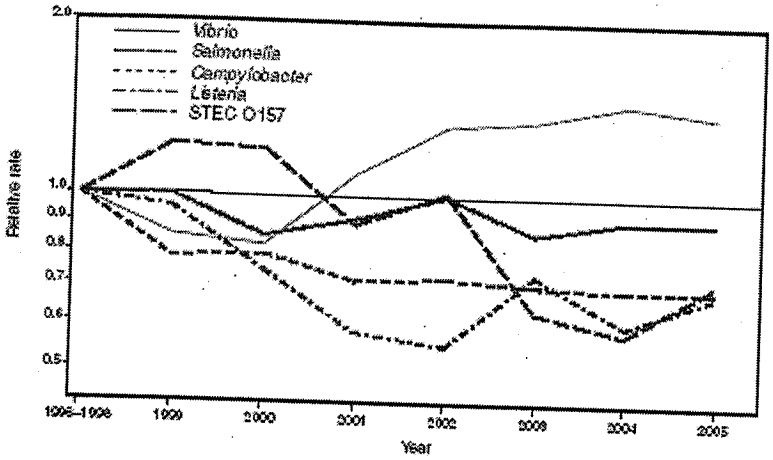
References

1. Allos BM, Moore MR, Griffin PM, Tauxe RV. Surveillance for sporadic foodborne disease in the 21st century: the FoodNet perspective. Clin Infect Dis 2004;38(Suppl 3):S115--20.
2. Hardnett FP, Hoekstra RM, Kennedy M, Charles L, Angulo FJ; Emerging Infections Program FoodNet Working Group. Epidemiologic issues in study design and data analysis related to FoodNet activities. Clin Infect Dis 2004;38(Suppl 3):S121--6.
3. US Department of Health and Human Services. Healthy people 2010 (conference ed, in 2 vols). Washington, DC: US Department of Health and Human Services; 2000.
4. US Department of Agriculture, Food Safety and Inspection Service. Progress report on *Salmonella* testing of raw meat and poultry products, 1998--2005. Washington, DC: US Department of Agriculture; 2006. Available at http://www.fsis.usda.gov/science/progress_report_salmonella_testing/index.asp.
5. US Department of Agriculture, Food Safety and Inspection Service. *Salmonella* verification sample result reporting: agency policy and use in public health protection. Fed Regist 2006;71:9772--7. Available at http://www.fsis.usda.gov/OPPDE/rdad/FRPubs/04-026N.pdf.
6. Agasan A, Kornblum J, Williams G, et al. Profile of *Salmonella enterica* subsp. *enterica* (subspecies I) serotype 4,5,12:i:- strains causing food-borne infections in New York City. J Clin Microbiol 2002;40:1924--9.
7. Voetsch AC, Angulo FJ, Rabatsky-Ehr T, et al. Laboratory practices for stool-specimen culture for bacterial pathogens, including *Escherichia coli* O157:H7, in the FoodNet sites, 1995--2000. Clin Infect Dis 2004;38(Suppl 3):S190--7.
8. Naugle AL, Holt KG, Levine P, Eckel R. Food Safety and Inspection Service regulatory testing program for *Escherichia coli* O157:H7 in raw ground beef. J Food Prot 2005;68:462--8.
9. Naugle AL, Holt KG, Levine P, Eckel R. Sustained decrease in the rate of *Escherichia coli* O157:H7-positive raw ground beef samples tested by the Food Safety and Inspection Service. J Food Prot 2006;69:480--1.

* Connecticut, Georgia, Maryland, Minnesota, New Mexico, Oregon, Tennessee, and selected counties in California, Colorado, and New York.

Figure 1

FIGURE 1. Relative rates compared with 1996–1998 baseline period of laboratory-diagnosed cases of infection with *Campylobacter*, STEC* O157, *Listeria*, *Salmonella*, and *Vibrio*, by year — Foodborne Diseases Active Surveillance Network, United States, 1996–2005

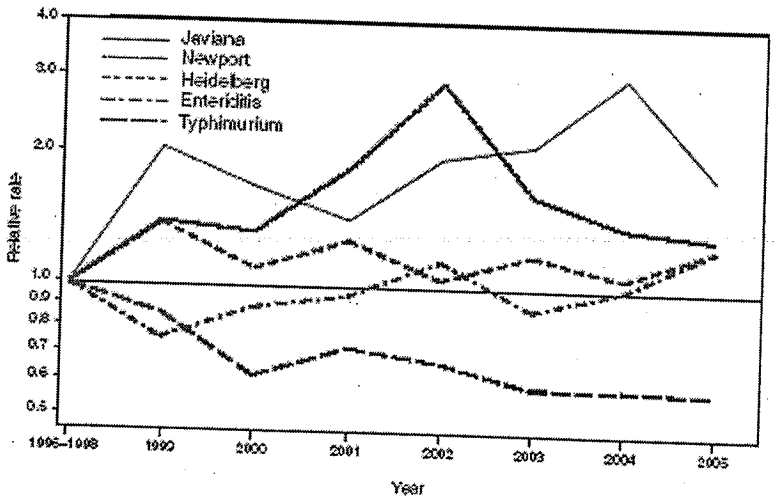


*Shiga toxin–producing *Escherichia coli*.

Return to top.

Figure 2

FIGURE 2. Relative rates compared with 1996–1998 baseline period of laboratory-diagnosed cases of infection with the five most commonly isolated *Salmonella* serotypes, by year — Foodborne Diseases Active Surveillance Network, United States, 1996–2005



Return to top.

Table

TABLE. Incidence* of cases of bacterial and parasitic infection and postdiarrheal hemolytic uremic syndrome (HUS), by site and pathogen, compared with national health objectives† — Foodborne Diseases Active Surveillance Network, United States, 2005§

Pathogen	California	Colorado	Connecticut	Georgia	Maryland	Minnesota	New Mexico	New York	Oregon	Tennessee	Overall 2005	National health objective†
Bacteria												
Campylobacter	27.96	19.37	15.47	6.52	7.23	16.51	18.28	11.70	17.69	6.98	12.72	12.30
Listeria	0.31	0.08	0.57	0.28	0.34	0.29	0.21	0.42	0.31	0.19	0.30	0.25
Salmonella	13.99	13.30	13.36	21.75	14.14	11.33	13.45	11.29	10.46	13.74	14.55	6.80
Shigella	8.70	3.95	1.66	7.48	1.78	1.90	6.94	1.53	2.36	8.49	4.67	N/A¶
STEC** O157	0.87	1.02	1.23	0.39	0.47	2.35	0.53	1.71	1.84	0.78	1.06	1.00
STEC non-O157	0.16	0.12	0.57	0.09	0.68	0.80	0.53	0.25	0.22	0.03	0.33	N/A
Vibrio	0.69	0.31	0.34	0.24	0.49	0.12	0.05	0.19	0.25	0.08	0.27	N/A
Yersinia	0.87	0.27	0.43	0.29	0.13	0.35	0.11	0.51	0.45	0.31	0.36	N/A
Parasites												
Cryptosporidium	1.43	0.94	2.34	1.64	0.61	3.22	1.05	16.38	1.34	0.73	2.95	N/A
Cyclospora	0.06	0.00	1.00	0.15	0.05	0.00	0.21	0.02	0.11	0.05	0.15	N/A
HUS††	0.94	1.02	0.47	0.44	0.80	1.51	0.00	0.83	1.33	2.34	0.94§§	0.9
Surveillance population (millions)	3.21	2.56	3.50	8.83	5.56	5.10	1.90	4.32	3.59	5.90	44.47	—

* Per 100,000 population.
† *Healthy People 2010* objectives for incidence of *Campylobacter*, *Salmonella*, and Shiga toxin–producing *Escherichia coli* O157 infections for year 2010 and for incidence of *Listeria* infections for year 2005.
§ 2004 data reported for HUS incidence.
¶ Not applicable because no national health objective exists regarding infection with this pathogen.
** Shiga toxin–producing *Escherichia coli*.
†† Incidence rate of postdiarrheal HUS in children aged <5 years; rate calculation is based on surveillance population aged <5 years.

Return to top.

SAFER·HEALTHIER·PEOPLE™
Morbidity and Mortality Weekly Report
Centers for Disease Control and Prevention
1600 Clifton Rd, MailStop K-95, Atlanta, GA
30333, U.S.A

USA.gov Government Made Easy

Department of Health and Human Services

Prevention of Hemolytic Uremic Syndrome (HUS) Caused by Infection with Shiga Toxin-Producing Escherichia coli (STEC) with Monoclonal Antibody Therapy

June 19, 2002
6700B Rockledge Drive, Room 1205
Bethesda, Maryland

Background: Hemolytic-uremic syndrome (HUS) is characterized by thrombocytopenia, nonimmune hemolytic anemia and renal insufficiency. It occurs most frequently in young children after a prodromal period of bloody diarrhea. The principal cause of HUS is infection with Shiga toxin-producing *Escherichia coli* (particularly serotype O157:H7). The clinical features of HUS have been well described, but its pathophysiology is only partly elucidated and definitive treatment is not available. Approximately 73,000 cases of infection with O157:H7 are thought to occur each year in the United States. About 10 to 15 percent of infected children are at risk of developing HUS.

It is thought that the Shiga toxins (Verotoxins) produced by *E. coli* O157:H7 gain access to the circulation, injure renal glomerular endothelial cells during the early stages of *E. coli* O157:H7 infection, and initiate a pathophysiological cascade that leads to HUS. Studies in animal models have suggested that parenterally administered monoclonal antibodies (MAbs) with the capacity to neutralize the toxins produced by *E. coli* O157:H7 and other Shiga toxin-producing *E. coli* (STEC) strains are protective against systemic toxin-mediated disease. The development of humanized mouse monoclonal antibodies reactive with Shiga toxins (Stx) types 1 and 2 provides the potential for their use as therapeutic and/or prophylactic modalities in human illness. NIAID has collaborated in an effort to develop these MAbs as clinically useful products. The MAbs are now at the stage where the IND process could be initiated.

In order to get input about how the use of these MAbs and other similar products could affect the development of HUS, NIAID convened an Expert Panel. The panel was asked to address four questions as the foundation for the discussions. The responses to each of the questions, recommendations and conclusions are summarized below.

Question 1: ***Given what we know about the pathogenic mechanisms of infection with Shiga toxin-producing E. coli (STEC) and the subsequent development of HUS in some children, what is the likelihood that treatment of STEC-infected children with toxin-specific monoclonal (MAbs) could***

potentiate the development and/or severity of watery or bloody diarrhea or HUS?

Although there are a number of hypotheses about the pathophysiological basis of HUS, current understanding of the basis of disease is limited. Studies in humans are difficult and human biopsy material is available principally as retrospective patient specimens. Most nephrologists do not see HUS patients early enough in the course of their illness to enable study of early stage disease. Measurement of circulating factors and/or immune cells can provide some information, but the potential number of measurable factors and alterations in their levels makes such studies complex.

The average seropositivity (by IgG antibody) in an urban population ranging in age from 0-60 years was found to be about 45% to Stx2 and about 12% to Stx1, suggesting that exposure to STEC is more common than initially assumed. The fact that the most susceptible populations for HUS are the very young and the very old suggests that lack of immunity or waning immunity is a susceptibility factor and that natural exposure provides protection. Antibodies that block toxins have been successfully used in a number of human diseases. Data from animal model studies indicate that MAbs to Stx may be useful as therapeutic and prophylactic agents to STEC-associated disease.

From an overall immunopathologic perspective, there is no evidence (with the exception of one observation from the 1980s) that there are autoantibodies in HUS as occur in post-infectious types of pathology such as rheumatic fever. There is no evidence for HUS being an immune complex or serum sickness-like illness. There is limited, at best, evidence for complement-mediated injury to the host as the basis of HUS. However, some data are emerging regarding the coagulation system in HUS. First, the coagulation abnormalities in HUS differ from those observed in classic disseminated intravascular coagulation (DIC), because the concentration of fibrinogen is normal or elevated, and the prothrombin time and partial-thromboplastin time are normal or only slightly prolonged. However, the concentration of circulating prothrombin fragment 1+2 (the peptide that is cleaved from prothrombin when thrombin is generated) increases. Elevated plasminogen-activator inhibitor type 1 (PAI-1) activity and increased tissue plasminogen activator (t-PA) antigen and D-dimer concentrations in plasma further characterize the coagulopathy of STEC-induced HUS. These abnormalities precede the development of renal injury in infected children.

Among the hypotheses for HUS pathogenesis are those that propose an association with complement activation. There was considerable discussion about the possibilities and mechanisms whereby activation of the complement pathways might affect outcomes in patients infected with STECs and about whether treatment with MAbs to Stx might alter disease outcome. It was suggested that the diarrheal stage and HUS be considered separately in terms of

pathology and effects of the MAb treatment. Complement activation in the course of HUS should be considered separately from complement activation due to treatment. It was noted that there are microbial products that can activate complement in the absence of antibody. Whether this pertains to the pathophysiology of STEC-associated illness has yet to be shown. Measurement of complement levels alone is an indirect, and likely insufficient, approach to this issue.

The pathway of toxin trafficking in STEC infection and HUS remains unclear. Studies have shown that the STEC toxins are associated with white blood cells in the circulation; free toxin in the circulation has not been detected. Free toxin can be found in fecal filtrates. The toxins are found for a number of days suggesting that there might be an interval in which MAb can be given, before eukaryotic cell injury occurs. While it is not known whether the toxins can adhere to endothelial cell surfaces, and in that way lead to complement activation and initiation of procoagulant reactions, very little toxin has been demonstrated on the (renal) vessel endothelium or other glomerular tissue, suggesting that large amounts of Stx-antibody complexes might not be formed. Also, toxins are found associated with neutrophils in the blood of healthy family members of HUS cases. However, studies have shown that toxin present on granulocytes is cytotoxic for human glomerular endothelial cells.

Recent studies have suggested that prior to the overt development of HUS many of the pathophysiologic derangements observed during HUS are already manifest in laboratory abnormalities in many patients, including generation of thrombin, inhibition of fibrinolysis and degradation of von Willebrand factor multimers.

In vitro studies indicate that the Shiga toxins can damage cells. Transfer of toxin from neutrophils to glomerular endothelial cells is seen in *In vitro* studies. Anti-Stx antibodies could block this transfer. However, it is difficult to extrapolate these findings to in vivo clinical effects. It is not clear which are the target tissues in the kidney, and how best to minimize Stx-mediated injury to the host. Fc receptors on cells should also be considered for their potential role in immunopathology, especially if an Stx molecule is bound by a monoclonal antibody. However, it should be noted that immune complex deposition in the kidney is not generally observed during STEC-induced HUS.

There was discussion of the use of MAb as intervention modalities in model systems and in other diseases. MAbs have been successfully employed to treat various diseases. However, a theoretical possibility exists that MAb treatment might worsen the disease. It is also possible to modify the MAbs and their properties so as to make them more effective. It was also noted that there are many systems in which MAb don't activate complement. Each system of MAb and its antigen(s) has its own properties and effects.

There have been few (and no controlled) studies of treatment of HUS patients with intravenous pooled gamma globulin. In this very small number of patients, no additional pathology appeared to be caused by this treatment nor was benefit established. In this context, it was also noted that circulating immunoglobulin could suppress unwanted complement effects.

There are various animal models of HUS, all of which differ from human disease. Based on the studies to date of the MAbs in mouse models, protective effects have been demonstrated without indication of potentiation of disease pathology. In the mouse model, in which renal damage is principally tubular and not glomerular, antibody can prevent kidney damage and death even when given 48-72 hours post-infection. Even those animals that fail to survive seem to have less pathological effects than animals not receiving antibody. The mouse model would be amenable to additional studies of tubular effects.

Ferrets were discussed as a possible animal model that might allow for screening for deleterious effects of the MAb. The model relies on oral/enteral administration of STEC to induce disease. The ferrets develop glomerular lesions and thromocytopenia somewhat reminiscent of human HUS. The drawbacks of this model are that because only 20-25% of infected animals develop glomerular lesions, a large number of animals would need to be studied in order to have meaningful results. Additionally, multiple blood draws are necessary, which is technically difficult because of the ferret's small size. Studies with the Stx MAbs have not yet been done in this model. Unpublished observations suggest that more tissue damage can be detected in ferrets when electron microscopy rather than light microscopy is used for assessment. It is not known if ferrets develop coagulopathies and specific immunological reagents for the coagulation system of ferrets are not available, further complicating experiments in this model.

The use of the baboon model was discussed briefly. The unnatural (intravenous) route of toxin delivery in this model makes extrapolations to human HUS and the relevance of effects of MAb in this system uncertain. There is also a rabbit model in which Stx2c-producing strains cause a severe hemorrhagic colitis, but no nephropathy. Thus, no known animal model parallels human HUS after STEC infection.

The experience with the formalin-inactivated respiratory syncytial virus (RSV) vaccine, in which disease was potentiated in vaccinated children, forms the background of concern for intervention in childhood diseases and the basis for trying to predict the potential for deleterious treatment effects. It was pointed out that the product under discussion is a MAb and not a vaccine, and that MAb against RSV are used prophylactically in children without ill effect.

Although studies in the animal models could serve to demonstrate a lack of deleterious effects, the ultimate test of the MAbs would be their use in children.

Question 2: ***What are considered the most important characteristics of a MAb for treatment of STEC-infected children for the purpose of preventing HUS?***

The STEC MAbs are being formulated using methodologies similar to that of MAb products currently on the market. It was suggested that formulation of the product be similar to that of the RSV monoclonal antibody for premature infants. Formulation should recognize that in sick children treatment would be started when the patients have low protein levels. The proposed dosage should take into account the presence of vascular leakage in these children, and the resulting volume of distribution of circulating IgG.

The MAbs to Stx1 and Stx2 would be available in separate vials. It is anticipated that they will be given together with more Stx2 than Stx1 (about 3:1) in order to allow for neutralization of the variants of Stx2. (Coverage of Stx1, Stx2, Stx2c, and Stx2d activatable appeared to be the most clinically relevant.)

The proposed Phase I protocol will include measurements of human anti-chimeric antibody (HACA). With MAb products currently on the market, anti-idiotypic antibodies have not been a problem. Only one or two injections with MAbs are anticipated. The MAb are specific and do not seem to react with other bacterial antigens or with human tissues in the testing performed to date (not discussed at this meeting). The use of both Stx1 and Stx2 MAbs is appropriate given that both toxins can cause severe disease. It was also noted that early in the twentieth century, large amounts of horse anti-diphtheria antibodies had been given to many children without uniformly adverse effects, although there was some occurrence of serum sickness.

The discussion of usage of antibody fragments, rather than whole antibody, concluded that the whole antibody would pose fewer problems in terms of site and rate of clearance. (Fab complexes are more likely to be cleared more rapidly than whole antibody and cleared by the kidney rather than by the liver; however, the site of clearance may not be relevant.)

Question 3: ***What is the definition of the target population, i.e., child presenting with diarrhea caused by STEC?***

There are two populations at high-risk for HUS, young children and the elderly. Data from population-based studies measuring IgG to Stx1 and Stx2 in urban and rural areas show a consistent pattern of lower antibody levels in children and in the elderly, which correlates with higher risk of HUS in these populations. Several overlapping databases on HUS were presented including subject identification in emergency rooms, by stool microbiology, from contacts of those with diarrhea, and from outbreak situations. Most cases of HUS are sporadic rather than outbreak associated. The proportions of cases of diarrhea due to STECs and of these, those who developed HUS, are modest and this will impact

on the design of efficacy studies because of the need to study a very large number of cases prospectively and who will need to be identified early in illness to receive treatment. Given the disease rates, it is difficult to ascertain if there have been trends or changes in incidence. The number of clinical laboratories that test for O157:H7 is only about 60 percent. (There are recent decreases in Campylobacter and Salmonella incidence and so it is possible that O157:H7 could be diminishing, but the current year data are needed to assess this.) In the Pacific Northwest/Seattle area, the number of severe outcomes such as seizures, strokes and fatalities has diminished since 1993. This may be a consequence of the appreciation by clinicians in the Pacific Northwest/Seattle area of the clinical significance of bloody diarrhea and the rapid administration of intravenous fluids to patients presenting with such symptoms.

There was substantial discussion about defining populations appropriate for use of the chimeric MAb. It was noted that it is difficult to define the point for optimal benefit, as intervention in the sickest children could be at a point where the MAb might be less effective or there might be a greater (theoretical) risk of potentiating illness. It appeared that children less than 10 years of age, with bloody diarrhea, brought to the hospital/ER, and having clinical or bacteriological evidence of an *E. coli* O157:H7 infection, could hypothetically benefit from MAb as a therapeutic, as up to 20 percent of these children go on to develop HUS. However, if bloody diarrhea in an Emergency Department setting is the criterion, a much lower percent of subjects will be at risk of developing HUS (probably under 5%), because of inclusion of patients with diarrhea due to pathogens with no or minimal risk of causing HUS.

An appropriate immunoprophylactic usage of the MAbs could be in an institutional outbreak (e.g., day care, nursing home) or when a public-health agency identified point-source exposure (e.g., food-borne outbreak). In this latter situation, the young and elderly, rather than all exposed persons, would be the target groups.

Reliable rapid diagnostic tools for STEC are not currently available, but are being developed. The exact age for inclusion criteria for children might be adjusted depending upon local epidemiological information. For outbreak situations, "ready-to-go" protocols would need to be available for prophylactic use of the MAb during an ongoing epidemiological investigation. In view of the kinetics of outbreak detection, most cases that will subsequently develop HUS, or become symptomatic, are already having symptoms by the time the "at risk" group comes to light. While more people are infected during an outbreak, they often have milder disease than would be seen in a hospital/ER situation evaluating sporadic illness. It should be emphasized that therapies applied to subjects in an outbreak might not be generalizable to all strains of *E. coli* O157:H7.

If toxin is playing a major role in the pathology of HUS and continuing to be present in the gut, on neutrophils, and in the circulation, the administration of the

MAb could ameliorate the disease process if binding to the MAb sequesters toxin away from target sites. On the other hand, there was some concern that without a better understanding of the pathophysiology of disease, perhaps the opposite could occur and disease potentiation would result.

Randomized and blinded studies would be needed to demonstrate therapeutic or prophylactic benefits, but the design of the trial and study protocol were not part of the charge to the expert panel and as such were not discussed in great detail.

Question 4: ***Do the existing preclinical data justify moving the MAb therapy forward into clinical studies in a) Phase I - healthy adults, and b) Phase II - sick children? Please evaluate from the standpoint of both safety and efficacy.***

The proposed Phase I toxicity studies of the chimeric monoclonal antibodies in healthy adults were briefly described to the panel. They appear to be appropriate for consideration for implementation. The current study design of intravenous administration was selected based on the preferred route for therapeutic administration. If the MAbs are eventually used as prophylactic agents, future evaluation of intramuscular administration could be studied as data for that indication are developed. Studies of complement fixation in healthy individuals would be difficult as no antigen (Stx) will be present in probands of the Phase I studies and any changes in complement levels would be quite low as the MAb would be in a "sea of complement." However, there are two circumstances in which one might consider such studies. If the antibody were altered during manufacture or if it encountered an unexpected antigen to which it reacted, measurable changes might occur. Thus, if the MAb disappears from the circulation very rapidly, one might consider doing complement studies, e.g., C4 levels.

Given the current lack of understanding of the pathophysiology of HUS, it may be more appropriate to ascertain the pharmacokinetics of the MAb in healthy children and in sick children without HUS. At this point in time, MAb is not appropriate for compassionate therapeutic use.

The risk for complications in prophylactic studies in children exposed to STEC infections is probably much less than in studies in sick children. However, a substantially larger cohort will be needed to achieve comparable statistical power. Obtaining truly informed consent for studies in sick children will be very difficult and will complicate the logistics of the clinical studies.

General Discussion:

It would be important to understand the normal immune response to the Stx toxins, in terms of immunoglobulin classes and IgG subtypes, and the steps involved between the introduction of toxins into the body and the induction of the

coagulopathies and other pathological effects. This might provide insight about the potential for immunopotentiation by monoclonal antibodies or reassurance that this was not likely.

While the historical observation of disease potentiation with formalinized RSV vaccine provides pause, there is some current experience with clinical use of MAb that suggests that there could be merit to this approach. A monoclonal RSV antibody is currently being used as a prophylactic modality for high-risk children. Although the RSV monoclonal had not been effective when used as a therapeutic modality, it is effective in preventing RSV disease. Monoclonal antibodies to Staphylococcus are currently in Phase I trials in adults and neonates. In addition, polyclonal antibodies to toxin-mediated diseases have a long history of successful medical use.

The long-term consequences of HUS are not known. There is some suggestion that there could be increased risk of kidney problems and hypertension. However, given the interval between HUS and the later development of clinical problems, plausibly related to HUS which may also be affected by other variables (such as smoking, and over-weight) that could impact on pathological consequences; this would be very difficult to study and verify.

It was also noted that STEC have been discussed as potential bioterrorism agents and the monoclonal antibodies, although not initially conceived for this purpose, could have relevance in the biodefense context.

Conclusions:

There is clearly a need to develop therapeutic, and preferably prophylactic, modalities for HUS. Phase I pharmacokinetic studies in adults with intravenous administration of escalating doses of anti-Stx MAb are appropriate at this time with the assumption that the long-term goal is to proceed with studies in at-risk populations. Before proceeding with studies into other phases and other groups, it is important to have more data on the pathophysiology of HUS to determine its underlying basis, and thus be able to better assess the risk for MAb to diminish, rather than enhance disease. Additional *In vitro* and animal model studies may provide more insights into pathogenesis and provide safety assurance for further studies in humans. Characterization of the normal immune response to toxin in naturally occurring infections in both asymptomatic and symptomatic humans could provide important insights into the potential for MAb to parallel such protective antibody responses rather than potentiate pathological responses. While animal studies can provide additional information, as in other diseases, only studies in humans can allow for the determination of the efficacy of the MAbs in treating and/or preventing HUS.

EXPERT PANEL MEMBERS
Prevention of Hemolytic Uremic Syndrome (HUS) Caused by Infection with

Shiga Toxin-Producing Escherichia coli (STEC) with Monoclonal Antibody Therapy
June 19, 2002

Invited Expert	Area of Expertise
David Acheson, M.D. Associate Professor University of Maryland School of Medicine Department of Epidemiology and Preventive Medicine	Research in enteric diseases
Martin M. Bitzan, M.D. Department of Pediatrics Pediatric Nephrology Wake Forest University Baptist Medical Center	Pediatric nephrologist Infectious diseases Physician Scientist Endothelial/ renal tubule cell injury and HUS
Michael Frank, M.D. Professor, and Chair Department of Pediatrics Children's Health Center Durham, NC	Immunology Expertise Complement
Patricia Griffin, M.D. MS A38 Foodborne Diseases Epidemiology Section DBMD, NCID, CDC	Epidemiologist
Mohamed Karmali, M.D. Director General Laboratory for Foodborne Zoonoses Health Canada	Medical microbiology Infectious diseases
Alison D. O'Brien, Ph.D. Chair, Department of Microbiology and Immunology Uniformed Services University of the Health Sciences	Microbiology Developed Shiga toxin MAb Animal models of HUS
Marguerite Neill, M.D. Brown University School of Medicine Memorial Hospital, Division of Infectious	Infectious Diseases Public Health Laboratory research

Disease	Epidemiology Clinical Trials
Phillip I. Tarr, M.D. Children's Hospital and Regional Medical Center Department of Pediatrics Seattle, WA	Pediatric Gastroenterologist Clinical and population-based research on diarrheagenic *E. coli*
Mark Taylor, M.D.* Department of Nephrology Birmingham Children's Hospital Kadywood Middleway United Kingdom	Nephrologist Renal immunobiology Rat model Cytokine responses
N.C.A.J. van de Kar, M.D. Dept. Pediatric Nephrology University Hospital Nijmegen The Netherlands	Pediatric Nephrologist
* Dr. Taylor was unable to attend, but provided written comments to the questions distributed to the panel members; his comments were included in the panel discussions.	

AND RADIOLOGICAL HEALTH for general information, or arrow down for specific topics.

Dated: October 7, 1997.

Joseph A. Levitt,

Deputy Director for Regulations Policy, Center for Devices and Radiological Health.

[FR Doc. 97-32875 Filed 12-16-97; 8:45 am]

BILLING CODE 4160-01-F

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration

Cardiovascular and Renal Drugs Advisory Committee; Notice of Meeting

AGENCY: Food and Drug Administration, HHS.

ACTION: Notice.

This notice announces a forthcoming meeting of a public advisory committee of the Food and Drug Administration (FDA). The meeting will be open to the public.

Name of Committee: Cardiovascular and Renal Drugs Advisory Committee.

General Function of the Committee: To provide advice and recommendations to the agency on FDA regulatory issues.

Date and Time: The meeting will be held on January 27, 1998, 8:30 a.m. to 5:30 p.m.; and January 28, 1998, 9 a.m. to 4 p.m.

Location: National Institutes of Health, Natcher Conference Center, 45 Center Dr., Bethesda, MD 20892.

Contact Person: Joan C. Standaert, Center for Drug Evaluation and Research (HFD-110), 419-259-6211, or Danyiel A. D'Antonio (HFD-21), 301-443-5455, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, or FDA Advisory Committee Information Line, 1-800-741-8138 (301-443-0572 in the Washington, DC area), code 12533. Please call the Information Line for up-to-date information on this meeting.

Agenda: On January 27, 1998, the committee will review and discuss: (1) New drug application (NDA) 20-736, Verdia™ (tasosartan, Wyeth-Ayerst Research), as a therapy for hypertension; and (2) the unapproved outpatient use of intermittent intravenous positive inotropic agents. On January 28, 1998, the committee will review and discuss NDA 20-718, Integrilin™ (eptifibatide, Cor Therapeutics, Inc.), for use in the settings of percutaneous transluminal angioplasty and acute coronary syndrome.

Procedure: Interested persons may present data, information, or views, orally or in writing, on issues pending before the committee. Written submissions may be made to the contact person by January 20, 1998. Oral presentations from the public will be scheduled between approximately 8:30 a.m. and 9:30 a.m. on January 27, 1998. Time allotted for each presentation may be limited. Those desiring to make formal oral presentations should notify the contact person before January 20, 1998, and submit a brief statement of the general nature of the evidence or arguments they wish to present, the names and addresses of proposed participants, and an indication of the approximate time requested to make their presentation.

Notice of this meeting is given under the Federal Advisory Committee Act (5 U.S.C. app. 2).

Dated: December 11, 1997.

Michael A. Friedman,

Deputy Commissioner for Operations.

[FR Doc. 97-32874 Filed 12-16-97; 8:45 am]

BILLING CODE 4160-01-F

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration

[Docket No. 97D-0188]

International Conference on Harmonisation; Guidance on General Considerations for Clinical Trials

AGENCY: Food and Drug Administration, HHS.

ACTION: Notice.

SUMMARY: The Food and Drug Administration (FDA) is publishing a guidance entitled "E8 General Considerations for Clinical Trials." The guidance was prepared under the auspices of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). The guidance sets forth general scientific principles for the conduct, performance, and control of clinical trials.

DATES: Effective December 17, 1997. Submit written comments at any time.

ADDRESSES: Submit written comments on the guidance to the Dockets Management Branch (HFA-305), Food and Drug Administration, 12420 Parklawn Dr., rm. 1-23, Rockville, MD 20857. Copies of the guidance are available from the Drug Information Branch (HFD-210), Center for Drug Evaluation and Research, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, 301-827-4573. Single copies of the guidance may be obtained by mail from the Office of Communication, Training and Manufacturers Assistance (HFM-40), Center for Biologics Evaluation and Research (CBER), Food and Drug Administration, 1401 Rockville Pike, Rockville, MD 20852-1448, or by calling the CBER Voice Information System at 1-800-835-4709 or 301-827-1800. Copies may be obtained from CBER's FAX Information System at 1-888-CBER-FAX or 301-827-3844.

FOR FURTHER INFORMATION CONTACT:

Regarding the guidance: G. Alexander Fleming, Center for Drug Evaluation and Research (HFD-510), Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, 301-827-6391.

Regarding ICH: Janet J. Showalter, Office of Health Affairs (HFY-20), Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, 301-827-0864.

SUPPLEMENTARY INFORMATION: In recent years, many important initiatives have been undertaken by regulatory authorities and industry associations to promote international harmonization of regulatory requirements. FDA has participated in many meetings designed to enhance harmonization and is committed to seeking scientifically based harmonized technical procedures for pharmaceutical development. One of the goals of harmonization is to identify and then reduce differences in technical requirements for drug development among regulatory agencies.

ICH was organized to provide an opportunity for tripartite harmonization initiatives to be developed with input from both regulatory and industry representatives. FDA also seeks input from consumer representatives and others. ICH is concerned with harmonization of technical requirements for the registration of pharmaceutical products among three regions: The European Union, Japan, and the United States. The six ICH sponsors are the European Commission, the European Federation of Pharmaceutical Industries Associations, the Japanese Ministry of Health and Welfare, the Japanese Pharmaceutical Manufacturers Association, the Centers for Drug Evaluation and Research (CDER) and CBER, FDA, and the Pharmaceutical Research and Manufacturers of America. The ICH Secretariat, which coordinates the preparation of documentation, is provided by the International Federation of Pharmaceutical Manufacturers Associations (IFPMA).

The ICH Steering Committee includes representatives from each of the ICH sponsors and the IFPMA, as well as observers from the World Health Organization, the Canadian Health Protection Branch, and the European Free Trade Area.

In the **Federal Register** of May 30, 1997 (62 FR 29540), FDA published a draft tripartite guideline entitled "General Considerations for Clinical Trials." The notice gave interested persons an opportunity to submit comments by July 1, 1997.

After consideration of the comments received and revisions to the guidance, a final draft of the guidance was submitted to the ICH Steering Committee and endorsed by the three participating regulatory agencies on July 17, 1997.

In accordance with FDA's Good Guidance Practices (62 FR 8961, February 27, 1997), this document has been designated a guidance, rather than a guideline.

The guidance describes internationally accepted principles and practices in the conduct of clinical trials and development strategy for new drug products. It is intended to facilitate the evaluation and acceptance of foreign clinical trial data by promoting a common understanding of general principles and approaches. The guidance also presents an overview of ICH clinical safety and efficacy documents.

This guidance represents the agency's current thinking on general considerations for the conduct, performance, and control of clinical trials. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

The public is encouraged to submit written comments with new data or other new information pertinent to this guidance. The comments in the docket will be periodically reviewed, and, where appropriate, the guidance will be amended. The public will be notified of any such amendments through a notice in the **Federal Register**.

Interested persons may, at any time, submit written comments on the guidance to the Dockets Management Branch (address above). Two copies of any comments are to be submitted, except that individuals may submit one copy. Comments are to be identified with the docket number found in brackets in the heading of this document. The guidance and received comments may be seen in the office above between 9 a.m. and 4 p.m., Monday through Friday. An electronic version of this guidance is available on the Internet (http://www.fda.gov/cder/guidance/index.htm) or at CBER's World Wide Web site at "http://www.fda.gov/cber/publications.htm".

The text of the guidance follows:

E8 General Considerations for Clinical Trials[1]

1. Objectives of This Document

In the three ICH regions, the evolution of drug development strategies and evaluation processes has led to the establishment of regional guidances on general considerations for clinical trials and the process of clinical development of pharmaceuticals for human use. This harmonized guidance is derived from those regional documents as well as from ICH guidances.

The ICH document "General Considerations for Clinical Trials" is intended to:

(a) Describe internationally accepted principles and practices in the conduct of both individual clinical trials and overall development strategy for new medicinal products.

(b) Facilitate the evaluation and acceptance of foreign clinical trial data by promoting a common understanding of general principles, general approaches, and the definition of relevant terms.

(c) Present an overview of the ICH clinical safety and efficacy documents and facilitate the user's access to guidance pertinent to clinical trials within these documents. The relevant ICH documents are listed in Annex 1.

(d) Provide a separate glossary of terms used in the ICH clinical safety and efficacy related documents that pertain to clinical trials and indicate which documents contain these.

For the sake of brevity, the term "drug" has been used in this document. It should be considered synonymous with "investigational (medicinal) product," "medicinal product," and "pharmaceutical," including vaccines and other biological products. The principles established in this guidance may also be applied to other clinical investigations (e.g., radiotherapy, psychotherapy, surgery, medical devices and alternative therapies).

2. General Principles

2.1 Protection of Clinical Trial Subjects

The principles and practices concerning protection of trial subjects are stated in the ICH guidance on Good Clinical Practice (ICH E6). These principles have their origins in The Declaration of Helsinki and should be observed in the conduct of all human drug investigations.

Before any clinical trial is carried out, results of nonclinical investigations or previous human studies should be sufficient to indicate that the drug is acceptably safe for the proposed investigation in humans. The purpose and timing of animal pharmacology and toxicology studies intended to support studies of a given duration are discussed in ICH M3. The role of such studies for biotechnology products is cited in ICH S6.

Throughout drug development, emerging animal toxicological and clinical data should be reviewed and evaluated by qualified experts to assess their implications for the safety of the trial subjects. In response to such findings, future studies and, when necessary, those in progress should be appropriately modified in a timely fashion to maintain the safety of trial participants. The investigator and sponsor share responsibility for the protection of clinical trial subjects together with the Institutional Review Board/Independent Ethics Committee. The responsibilities of these parties are described in ICH E6.

2.2 Scientific Approach in Design and Analysis

Clinical trials should be designed, conducted, and analyzed according to sound scientific principles to achieve their objectives, and should be reported appropriately. The essence of rational drug development is to ask important questions and answer them with appropriate studies. The primary objectives of any study should be clear and explicitly stated.

Clinical studies can be classified according to when the study occurs during clinical development or, as shown in Table 1, by their objectives. (The illustrative examples are not intended to be exhaustive.) The cardinal logic behind serially conducted studies of a medicinal product is that the results of prior studies should influence the plan of later studies. Emerging data will frequently prompt a modification of the development strategy. For example, results of a therapeutic confirmatory study may suggest a need for additional human pharmacology studies.

The availability of foreign clinical data should obviate the need to generate similar data in an ICH region if the ICH E5 and ICH E6 guidances are followed (see ICH E5).

[1] This guidance represents the agency's current thinking on general considerations for the conduct, performance, and control of clinical trials. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

TABLE 1.—AN APPROACH TO CLASSIFYING CLINICAL STUDIES ACCORDING TO OBJECTIVE

Type of Study	Objective of Study	Study Examples
Human Pharmacology	• Assess tolerance • Define/describe PK[1] and PD[2] • Explore drug metabolism and drug interactions	• Dose-tolerance studies • Single and multiple dose PK and/or PD studies • Drug interaction studies
Therapeutic Exploratory	• Estimate activity • Explore use for the targeted indication • Estimate dosage for subsequent studies • Provide basis for confirmatory study design, endpoints, methodologies	• Earliest trials of relatively short duration in well-defined narrow patient populations, using surrogate or pharmacological endpoints or clinical measures • Dose-response exploration studies
Therapeutic Confirmatory	• Demonstrate/confirm efficacy • Establish safety profile • Provide an adequate basis for assessing the benefit/risk relationship to support licensing • Establish dose-response relationship	• Adequate, and well controlled studies to establish efficacy • Randomized parallel dose-response studies • Clinical safety studies • Studies of mortality/morbidity outcomes • Large simple trials
Therapeutic Use	• Refine understanding of benefit/risk relationship in general or special populations and/or environments • Identify less common adverse reactions • Refine dosing recommendation	• Comparative studies • Comparative effectiveness studies • Studies of mortality/morbidity outcomes • Studies of additional endpoints • Large simple trials • Pharmacoeconomic studies

[1] Pharmacokinetics
[2] Pharmacodynamics

3. Development Methodology

This section covers issues and considerations relating to the development plan and to its individual component studies.

3.1 Considerations for the Development Plan

3.1.1 Nonclinical Studies

Important considerations for determining the nature of nonclinical studies and their timing with respect to clinical trials include:

(a) Duration and total exposure proposed in individual patients.

(b) Characteristics of the drug (e.g., long half life, biotechnology products).

(c) Disease or condition targeted for treatment.

(d) Use in special populations (e.g., women of childbearing potential).

(e) Route of administration.

The need for nonclinical information including toxicology, pharmacology, and pharmacokinetics to support clinical trials is addressed in the ICH M3 and S6 documents.

3.1.1.1 *Safety studies.* For the first studies in humans, the dose that is administered should be determined by careful examination of the prerequisite nonclinical pharmacokinetic, pharmacological, and toxicological evaluations (see ICH M3). Early nonclinical studies should provide sufficient information to support selection of the initial human dose and safe duration of exposure, and to provide information about physiological and toxicological effects of a new drug.

3.1.1.2 *Pharmacological and pharmacokinetic studies.* The basis and direction of the clinical exploration and development rests on the nonclinical pharmacokinetic and pharmacology profile, which includes information such as:

(a) Pharmacological basis of principal effects (mechanism of action).

(b) Dose-response or concentration-response relationships and duration of action.

(c) Study of the potential clinical routes of administration.

(d) Systemic general pharmacology, including pharmacological effects on major organ systems and physiological responses.

(e) Studies of absorption, distribution, metabolism, and excretion.

3.1.2 Quality of Investigational Medicinal Products

Formulations used in clinical trials should be well characterized, including information on bioavailability wherever feasible. The formulation should be appropriate for the stage of drug development. Ideally, the supply of a formulation will be adequate to allow testing in a series of studies that examine a range of doses. During drug development, different formulations of a drug may be tested. Links between formulations, established by bioequivalence studies or other means, are important in interpreting clinical study results across the development program.

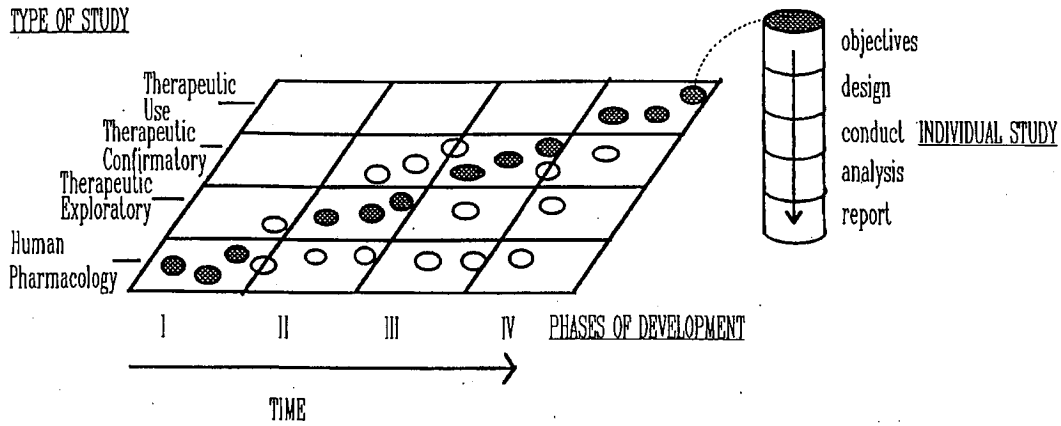
3.1.3 Phases of Clinical Development

Clinical drug development is often described as consisting of four temporal phases (Phases I–IV). It is important to recognize that the phase of development provides an inadequate basis for classification of clinical trials because one type of trial may occur in several phases (see Figure 1). A classification system using study objectives as discussed in section 2.2 is preferable. It is important to appreciate that the phase concept is a description, not a set of requirements. It is also important to realize that the temporal phases do not imply a fixed order of studies since for some drugs in a development plan the typical sequence will not be appropriate or necessary. For example, although human pharmacology studies are typically conducted during Phase I, many such studies are conducted at each of the other three stages, but nonetheless are sometimes labeled as Phase I studies. Figure 1 demonstrates this close but variable correlation between the two classification systems. The distribution of the points of the graph shows that the types of study are not synonymous with the phases of development.

BILLING CODE 4160–01–F

[1] This matrix graph illustrates the relationship between the phases of development and types of study by objective that may be conducted during each clinical development of a new medicinal product. The shaded circles show the types of study most usually conducted in a certain phase of development, the open circles show certain types of study that may be conducted in that phase of development but are less usual. Each circle represents an individual study. To illustrate the development of a single study, one circle is joined by a dotted line to an inset column that depicts the elements and sequence of an individual study.

Figure 1.—Correlation Between Development Phases and Types of Study[1]



BILLING CODE 4160-01-C

Drug development is ideally a logical, step-wise procedure in which information from small early studies is used to support and plan later larger, more definitive studies. To develop new drugs efficiently, it is essential to identify characteristics of the investigational medicine in the early stages of development and to plan an appropriate development based on this profile.

Initial trials provide an early evaluation of short-term safety and tolerability and can provide pharmacodynamic and pharmacokinetic information needed to choose a suitable dosage range and administration schedule for initial exploratory therapeutic trials. Later confirmatory studies are generally larger and longer and include a more diverse patient population. Dose-response information should be obtained at all stages of development, from early tolerance studies, to studies of short-term pharmacodynamic effect, to large efficacy studies (see ICH E4). Throughout development, new data may suggest the need for additional studies that are typically part of an earlier phase. For example, blood level data in a late trial may suggest a need for a drug-drug interaction study, or adverse effects may suggest the need for further dose finding and/or additional nonclinical studies. In addition, to support a new marketing application approval for the same drug, e.g., for a new indication, pharmacokinetic or therapeutic exploratory studies are considered to be in Phase I or Phase II of development.

3.1.3.1 *Phase I (Most typical kind of study: Human pharmacology).* Phase I starts with the initial administration of an investigational new drug into humans.

Although human pharmacology studies are typically identified with Phase I, they may also be indicated at other points in the development sequence. Studies in this phase of development usually have nontherapeutic objectives and may be conducted in healthy volunteer subjects or certain types of patients, e.g., patients with mild hypertension. Drugs with significant potential toxicity, e.g., cytotoxic drugs, are usually studied in patients. Studies in this phase can be open, baseline controlled, or may use randomization and blinding, to improve the validity of observations.

Studies conducted in Phase I typically involve one or a combination of the following aspects:

(a) Estimation of initial safety and tolerability

The initial and subsequent administration of an investigational new drug into humans is usually intended to determine the tolerability of the dose range expected to be needed for later clinical studies and to determine the nature of adverse reactions that can be expected. These studies typically include both single and multiple dose administration.

(b) Pharmacokinetics

Characterization of a drug's absorption, distribution, metabolism, and excretion continues throughout the development plan. Their preliminary characterization is an important goal of Phase I. Pharmacokinetics may be assessed via separate studies or as a part of efficacy, safety and tolerance studies. Pharmacokinetic studies are particularly important to assess the clearance of the drug and to anticipate possible accumulation of parent drug or metabolites and potential drug-drug interactions. Some pharmacokinetic studies are commonly conducted in later phases to answer more specialized questions. For many orally administered drugs, especially modified release products, the study of food effects on bioavailability is important. Obtaining pharmacokinetic information in subpopulations such as patients with impaired elimination (renal or hepatic failure), the elderly, children, women, and ethnic subgroups should be considered. Drug-drug interaction studies are important for many drugs; these are generally performed in phases beyond Phase I, but studies in animals and in vitro studies of metabolism and potential interactions may lead to doing such studies earlier.

(c) Assessment of pharmacodynamics

Depending on the drug and the endpoint studied, pharmacodynamic studies and studies relating drug blood levels to response (PK/PD studies) may be conducted in healthy volunteer subjects or in patients with the target disease. In patients, if there is an appropriate measure, pharmacodynamic data can provide early estimates of activity and potential efficacy and may guide the dosage and dose regimen in later studies.

(d) Early measurement of drug activity

Preliminary studies of activity or potential therapeutic benefit may be conducted in Phase I as a secondary objective. Such studies are generally performed in later phases but may be appropriate when drug activity is readily measurable with a short duration of drug exposure in patients at this early stage.

3.1.3.2 *Phase II (Most typical kind of study: Therapeutic exploratory).* Phase II is usually considered to start with the initiation of studies in which the primary objective is to explore therapeutic efficacy in patients.

Initial therapeutic exploratory studies may use a variety of study designs, including concurrent controls and comparisons with baseline status. Subsequent trials are usually randomized and concurrently controlled to evaluate the efficacy of the drug and its safety for a particular therapeutic indication. Studies in Phase II are typically conducted in a group of patients who are selected by relatively narrow criteria, leading to a relatively homogeneous population, and who are closely monitored.

An important goal for this phase is to determine the dose(s) and regimen for Phase III trials. Early studies in this phase often utilize dose escalation designs (see ICH E4) to give an early estimate of dose response and later studies may confirm the dose response relationship for the indication in question by using recognized parallel dose-response designs (could also be deferred to phase III). Confirmatory dose response studies may be conducted in Phase II or left for Phase III. Doses used in Phase II are usually but not always less than the highest doses used in Phase I.

Additional objectives of clinical trials conducted in Phase II may include evaluation of potential study endpoints, therapeutic regimens (including concomitant medications), and target populations (e.g., mild versus severe disease) for further study in Phase II or III. These objectives may be served by exploratory analyses, examining subsets of data, and by including multiple endpoints in trials.

3.1.3.3 *Phase III (Most typical kind of study: Therapeutic confirmatory).* Phase III usually is considered to begin with the initiation of studies in which the primary objective is to demonstrate, or confirm therapeutic benefit.

Studies in Phase III are designed to confirm the preliminary evidence accumulated in Phase II that a drug is safe and effective for use in the intended indication and recipient population. These studies are intended to provide an adequate basis for marketing approval. Studies in Phase III may also further explore the dose-response relationship, or explore the drug's use in wider populations, in different stages of disease, or in combination with another drug. For drugs intended to be administered for long periods, trials involving extended exposure to the drug are ordinarily conducted in Phase III, although they may be started in Phase II (see ICH E1). ICH E1 and ICH E7 describe the overall clinical safety database considerations for chronically administered drugs and drugs used in the elderly. These studies carried out in Phase III complete the information needed to support adequate instructions for use of the drug (official product information).

3.1.3.4 *Phase IV (Variety of studies: Therapeutic use).* Phase IV begins after drug approval. Therapeutic use studies go beyond the prior demonstration of the drug's safety, efficacy and dose definition.

Studies in Phase IV are all studies (other than routine surveillance) performed after drug approval and related to the approved indication. They are studies that were not considered necessary for approval but are often important for optimizing the drug's use. They may be of any type but should have valid scientific objectives. Commonly conducted studies include additional drug-drug interaction, dose-response, or safety studies and studies designed to support use under the approved indication, e.g., mortality/morbidity studies, epidemiological studies.

3.1.3.5 *Development of an application unrelated to original approved use.* After initial approval, drug development may continue with studies of new or modified indications, new dosage regimens, new routes of administration, or additional patient populations. If a new dose, formulation, or combination is studied, additional human pharmacology studies may be indicated, necessitating a new development plan.

The need for some studies may be obviated by the availability of data from the original development plan or from therapeutic use.

3.1.4 Special Considerations

A number of special circumstances and populations require consideration on their own when they are part of the development plan.

3.1.4.1 *Studies of drug metabolites.* Major active metabolite(s) should be identified and deserve detailed pharmacokinetic study. Timing of the metabolic assessment studies within the development plan depends on the characteristics of the individual drug.

3.1.4.2 *Drug-drug interactions.* If a potential for drug-drug interaction is suggested by metabolic profile, by the results of nonclinical studies or by information on similar drugs, studies on drug interaction during clinical development are highly recommended. For drugs that are frequently coadministered, it is usually important that drug-drug interaction studies be performed in nonclinical and, if appropriate, in human studies. This is particularly true for drugs that are known to alter the absorption or metabolism of other drugs (see ICH E7), or whose metabolism or excretion can be altered by effects of other drugs.

3.1.4.3 *Special populations.* Some groups in the general population may require special study because they have unique risk/benefit considerations that need to be taken into account during drug development, or because they can be anticipated to need modification of use of the dose or schedule of a drug compared to general adult use.

Pharmacokinetic studies in patients with renal and hepatic dysfunction are important to assess the impact of potentially altered drug metabolism or excretion. Other ICH documents address such issues for geriatric patients (ICH E7) and patients from different ethnic groups (ICH E5). The need for nonclinical safety studies to support human clinical trials in special populations is addressed in the ICH M3 document.

Particular attention should be paid to the ethical considerations related to informed consent from vulnerable populations and the procedures scrupulously followed (see ICH E6).

(a) Investigations in pregnant women

In general, pregnant women should be excluded from clinical trials where the drug is not intended for use in pregnancy. If a patient becomes pregnant during administration of the drug, treatment should generally be discontinued if this can be done safely. Followup evaluation of the pregnancy, fetus, and child is very important. Similarly, for clinical trials that include pregnant women because the medicinal product is intended for use during pregnancy, followup of the pregnancy, fetus, and child is very important.

(b) Investigations in nursing women

Excretion of the drug or its metabolites into human milk should be examined where applicable. When nursing mothers are enrolled in clinical studies, their babies should be monitored for the effects of the drug.

(c) Investigations in children

The extent of the studies needed depends on the current knowledge of the drug and the possibility of extrapolation from adults and children of other age groups. Some drugs may be used in children from the early stages of drug development (see ICH M3).

For a drug expected to be used in children, evaluation should be made in the appropriate age group. When clinical development is to include studies in children, it is usually appropriate to begin with older children before extending the trial to younger children and then infants.

3.2 Considerations for Individual Clinical Trials

The following important principles should be followed in planning the objectives, design, conduct, analysis, and reporting of a clinical trial (see ICH guidances in Annex 1). Each part should be defined in a written protocol before the study starts (see ICH E6).

3.2.1 Objectives

The objective(s) of the study should be clearly stated and may include exploratory or confirmatory characterization of safety and/or efficacy and/or assessment of pharmacokinetic parameters and pharmacological, physiological, or biochemical effects.

3.2.2 Design

The appropriate study design should be chosen to provide the desired information. Examples of study design include parallel group, crossover, factorial, dose escalation, and fixed dose-dose response (see ICH E4, E6, E9 and E10). Appropriate comparators should be utilized and adequate numbers of subjects included to achieve the study objectives. Primary and secondary endpoints and plans for their analyses should be clearly stated (see ICH E9). The methods of monitoring adverse events by changes in clinical signs and symptoms and laboratory studies should be described (see ICH E3). The protocol should specify procedures for the followup of patients who stop treatment prematurely.

3.2.2.1 *Selection of subjects.* The stage of development and the indication to be studied should be taken into account in selecting the subject population (e.g., normal healthy subjects, cancer patients or other special populations in early phase development) as should prior nonclinical and clinical knowledge. The variability of groups of patients or healthy volunteers studied in early trials may be limited to a narrow range by strict selection criteria, but as drug development proceeds, the populations tested should be broadened to reflect the target population.

Depending on the stage of development and level of concern for safety, it may be necessary to conduct studies in a closely monitored (i.e., inpatient) environment.

As a general principle, trial subjects should not participate concurrently in more than one clinical trial but there can be justified exceptions. Subjects should not be enrolled repetitively in clinical trials without time off treatment adequate to protect safety and exclude carryover effects.

In general, women of childbearing potential should be using highly effective contraception to participate in clinical trials (see ICH M3).

For male subjects, potential hazards of drug exposure in the trial to their sexual partners or resulting progeny should be considered. When indicated (e.g., trials involving drugs that are potentially mutagenic, or toxic to the reproductive system), an appropriate contraception provision should be included in the trial.

3.2.2.2 *Selection of control group.* Trials should have an adequate control group. Comparisons may be made with placebo, no treatment, active controls, or of different doses of the drug under investigation. The choice of the comparator depends on, among other things, the objective of the trial (see ICH E9 and E10). Historical (external) controls can be justified in some cases, but particular care is important to minimize the likelihood of erroneous inference.

3.2.2.3 *Number of subjects.* The size of a trial is influenced by the disease to be investigated, the objective of the study, and the study endpoints. Statistical assessments of sample size should be based on the expected magnitude of the treatment effect, the variability of the data, the specified (small) probability of error (see ICH E9), and the desire for information on subsets of the population or secondary endpoints. In some circumstances, a larger database may be needed to establish the safety of a drug. ICH E1 and ICH E7 suggest a minimum experience to assess safety for a registrational database for a new indication. These numbers should not be considered as absolute and may be insufficient in some cases (e.g., where long-term use in healthy individuals is expected).

3.2.2.4 *Response variables.* Response variables should be defined prospectively, giving descriptions of methods of observation and quantification. Objective methods of observation should be used where possible and when appropriate (see ICH E9).

Study endpoints are the response variables that are chosen to assess drug effects that are related to pharmacokinetic parameters, pharmacodynamic measures, efficacy and safety. A primary endpoint(s) should reflect clinically relevant effects and is typically selected based on the principal objective of the study. Secondary endpoints assess other drug effects that may or may not be related to the primary endpoint. Endpoints and the plan for their analysis should be prospectively specified in the protocol.

A surrogate endpoint is an endpoint that is intended to relate to a clinically important outcome but does not in itself measure a clinical benefit. Surrogate endpoints may be used as primary endpoints when appropriate (when the surrogate is reasonably likely or well known to predict clinical outcome).

The methods used to make the measurements of the endpoints, both subjective and objective, should be validated and meet appropriate standards for accuracy, precision, reproducibility, reliability, and responsiveness (sensitivity to change over time).

3.2.2.5 *Methods to minimize or assess bias.* The protocol should specify methods of allocation to treatment groups and blinding (see ICH E9 and E10).

(a) Randomization

In conducting a controlled trial, randomized allocation is the preferred means of assuring comparability of test groups and minimizing the possibility of selection bias.

(b) Blinding

Blinding is an important means of reducing or minimizing the risk of biased study outcomes. A trial where the treatment assignment is not known by the study participant because of the use of placebo or other methods of masking the intervention is referred to as a single blind study. When the investigator and sponsor staff who are involved in the treatment or clinical evaluation of the subjects and analysis of data are also unaware of the treatment assignments, the study is double blind.

(c) Compliance

Methods used to evaluate patient usage of the test drug should be specified in the protocol and the actual usage documented.

3.2.3 Conduct

The study should be conducted according to the principles described in this guidance and in accordance with other pertinent elements outlined in ICH E6 and other relevant ICH guidances (see Annex 1). Adherence to the study protocol is essential. If modification of the protocol becomes necessary, a clear description of the rationale for the modification should be provided in a protocol amendment (see ICH E6). Timely adverse event reporting during a study is essential and should be documented. Guidance is available on expedited reporting of safety data to appropriate officials, on the content of safety reports, and on privacy and confidentiality of data (see ICH E2A, E2B, and E6).

3.2.4 Analysis

The study protocol should have a specified analysis plan that is appropriate for the objectives and design of the study, taking into account the method of subject allocation, the measurement methods of response variables, specific hypotheses to be tested, and analytical approaches to common problems including early study withdrawal and protocol violations. A description of the statistical methods to be employed, including timing of any planned interim analysis(es), should be included in the protocol (see ICH E3, E6, and E9).

The results of a clinical trial should be analyzed in accordance with the plan prospectively stated in the protocol and all deviations from the plan should be indicated in the study report. Detailed guidance is available in other ICH guidances on planning of the protocol (ICH E6), on the analysis plan and statistical analysis of results (ICH E9), and on study reports (ICH E3).

Studies are normally expected to run to completion, although in some studies the possibility of early stopping is formally recognized. In such cases, this should be clearly described in the protocol with due statistical attention to the overall levels of statistical significance and to the need to adjust the estimates of the size of treatment effects (ICH E9).

Safety data should be collected for all clinical trials, appropriately tabulated and with adverse events classified according to their seriousness and their likely causal relationship (see ICH E2A).

3.2.5 Reporting

Clinical study reports should be adequately documented following the approaches outlined in other ICH guidances (see E3 and E6).

4. Annex 1

TABLE 2.—LIST OF RELEVANT ICH GUIDANCES AND TOPICS

Code	Topic
E1	The Extent of Population Exposure to Assess Clinical Safety for Drugs Intended for Long-Term Treatment of Non-Life-Threatening Conditions
E2A	Clinical Safety Data Management: Definitions and Standards for Expedited Reporting
E2B	Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports
E2C	Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs
E3	Structure and Content of Clinical Study Reports
E4	Dose-Response Information to Support Drug Registration
E5	Ethnic Factors in the Acceptability of Foreign Clinical Data
E6	Good Clinical Practice: Consolidated Guideline
E7	Studies in Support of Special Populations: Geriatrics
E8	General Considerations for Clinical Trials
E9	Statistical Considerations in the Design of Clinical Trials
E10	Choice of Control Group in Clinical Trials
M3	Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals
S6	Safety Studies for Biotechnology-Derived Products

Dated: December 10, 1997.

William K. Hubbard,

Associate Commissioner for Policy Coordination.

[FR Doc. 97-32877 Filed 12-16-97; 8:45 am]

BILLING CODE 4160-01-F

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Health Care Financing Administration

[Document Identifier: HCFA-R-26]

Agency Information Collection Activities: Proposed Collection; Comment Request

AGENCY: Health Care Financing Administration.

In compliance with the requirement of section 3506(c)(2)(A) of the Paperwork Reduction Act of 1995, the Health Care Financing Administration (HCFA), Department of Health and Human Services, is publishing the following summary of proposed collections for public comment. Interested persons are invited to send comments regarding this burden estimate or any other aspect of this collection of information, including any of the following subjects: (1) The necessity and utility of the proposed information collection for the proper performance of the agency's functions; (2) the accuracy of the estimated burden; (3) ways to enhance the quality, utility, and clarity of the information to be collected; and (4) the use of automated collection techniques or other forms of information technology to minimize the information collection burden.

Type of Information Collection Request:

Extension of a currently approved collection; *Title of Information Collection:* Clinical Laboratory Improvement Amendment (CLIA) and the Information Collection Requirements (ICRs) contained in the Supporting Regulations 42 CFR 493.1-2001; *Form No.:* HCFA-R-26 (OMB# 0938-0612); *Use:* The ICRs referenced in 42 CFR 493.1-.2001 outline the requirements necessary to determine an entities compliance with CLIA. CLIA requires laboratories that perform testing on human specimens to meet performance requirements in order to be certified by HHS. HHS conducts inspections in order to determine a laboratory's compliance with the CLIA requirements. CLIA implements certificate, laboratory standards and inspection requirements.; *Frequency:* As needed; *Affected Public:* Individuals or Households, Business or other for profit, Not for profit institutions, Federal Government, State, local or tribal government; *Number of Respondents:* 149,700; *Total Annual Responses:* 631,459; *Total Annual Hours:* 9,133,625.

To obtain copies of the supporting statement and any related forms for the proposed paperwork collections referenced above, E-mail your request, including your address, phone number, OMB number, and HCFA document identifier, to Paperwork@hcfa.gov, or call the Reports Clearance Office on (410) 786-1326. Written comments and recommendations for the proposed information collections must be mailed within 60 days of this notice directly to the HCFA Paperwork Clearance Officer designated at the following address: HCFA, Office of Information Services, Information Technology Investment Management Group, Division of HCFA Enterprise Standards, Attention: Louis Blank, Room C2-26-17, 7500 Security Boulevard, Baltimore, Maryland 21244-1850.

Dated: December 5, 1997.

John P. Burke III,

HCFA Reports Clearance Officer, HCFA Office of Information Services, Information Technology Investment Management Group, Division of HCFA Enterprise Standards.

[FR Doc. 97-32859 Filed 12-16-97; 8:45 am]

BILLING CODE 4120-03-P

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Health Care Financing Administration

Document Identifier: HCFA-R-205 and HCFA-R-206

Emergency Clearance: Public Information Collection Requirements Submitted to the Office of Management and Budget (OMB)

AGENCY: Health Care Financing Administration.

In compliance with the requirement of section 3506(c)(2)(A) of the Paperwork Reduction Act of 1995, the Health Care Financing Administration (HCFA), Department of Health and Human Services, is publishing the following summary of proposed collections for public comment. Interested persons are invited to send comments regarding this burden estimate or any other aspect of this collection of information, including any of the following subjects: (1) The necessity and utility of the proposed information collection for the proper performance of the agency's functions; (2) the accuracy of the estimated burden; (3) ways to enhance the quality, utility, and clarity of the information to be collected; and (4) the use of automated collection techniques or other forms of information technology to minimize the information collection burden.

We are, however, requesting an emergency review of the information collections referenced below. In compliance with the requirement of section 3506(c)(2)(A) of the Paperwork Reduction Act of 1995, we have submitted to the Office of Management and Budget (OMB) the following requirements for emergency review. We are requesting an emergency review because the collection of this information is needed before the expiration of the normal time limits under OMB's regulations at 5 CFR, Part 1320. This is necessary to ensure compliance with section 111 of HIPAA necessary to implement congressional intent with respect to guaranteeing availability of individual health insurance coverage to certain individuals with prior group coverage. We cannot reasonably comply with the normal clearance procedures because public harm is likely to result because eligible individuals will not receive the health insurance protections under the statute.

HCFA is requesting OMB review and approval of this collection by 12/31/97, with a 180-day approval period. Written comments and recommendations will be accepted from the public if received by the individuals designated below by 12/29/97. It should be noted that HCFA will continue to consider and respond as appropriate to the public comments received in response to the 04/08/97 **Federal Register** notices requesting public comment on the collections referenced below. During this 180-day period, we will publish a separate **Federal Register** notice announcing the initiation of an extensive 60-day agency review and public comment period on these requirements. We will submit the requirements for OMB review and an extension of this emergency approval.

Type of Information Request: Extension, without change, of a currently approved collection.

Title of Information Collection: Individual Health Insurance. Reform: Portability from Group to Individual Coverage; Federal Rules for Access in the Individual Market; State Alternative Mechanisms to Federal Rules BPD-882-IFC.

Form Number: HCFA-R-205 (OMB approval #: 0938-0703).

Use: These information collection requirements help ensure access to the individual insurance market for certain individuals and allows the States to

Guidance for Industry

E9 Statistical Principles for Clinical Trials

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
September 1998
ICH

Guidance for Industry

E9 Statistical Principles for Clinical Trials

Additional copies are available from:

Office of Training and Communications
Division of Drug Information (HFD-240)
Center for Drug Evaluation and Research (CDER),
5600 Fishers Lane, Rockville, MD 20857 (Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm

or

Office of Communication, Training, and Manufacturers Assistance (HFM-40)
Center for Biologics Evaluation and Research (CBER)
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm; (Fax) 888-CBERFAX or 301-827-3844
(Voice Information) 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
September 1998
ICH

TABLE OF CONTENTS

GUIDANCE FOR INDUSTRY[1]

E9 Statistical Principles for Clinical Trials

I. INTRODUCTION

A. Background and Purpose (1.1) [2]

The efficacy and safety of medicinal products should be demonstrated by clinical trials that follow the guidance in *E6 Good Clinical Practice: Consolidated Guidance* adopted by the ICH, May 1, 1996. The role of statistics in clinical trial design and analysis is acknowledged as essential in that ICH guidance. The proliferation of statistical research in the area of clinical trials coupled with the critical role of clinical research in the drug approval process and health care in general necessitate a succinct document on statistical issues related to clinical trials. This guidance is written primarily to attempt to harmonize the principles of statistical methodology applied to clinical trials for marketing applications submitted in Europe, Japan and the United States.

As a starting point, this guidance utilized the CPMP (Committee for Proprietary Medicinal Products) Note for Guidance entitled *Biostatistical Methodology in Clinical Trials in Applications for Marketing Authorizations for Medicinal Products* (December 1994). It was also influenced by *Guidelines on the Statistical Analysis of Clinical Studies* (March 1992) from the Japanese Ministry of Health and Welfare and the U.S. Food and Drug Administration document entitled *Guideline for the Format and Content of the Clinical*

---

[1] This guidance was developed within the Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, February 1998. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan, and the United States. This guidance was published in the *Federal Register* on September 16, 1998 (63 FR 49583), and is applicable to drug and biological products.

[2] Arabic numbers reflect the organizational breakdown in the document endorsed by the ICH Steering Committee at Step 4 of the ICH process, February 1998.

and Statistical Sections of a New Drug Application (July 1988). Some topics related to statistical principles and methodology are also embedded within other ICH guidances, particularly those listed below. The specific guidance that contains related text will be identified in various sections of this document.

E1A *The Extent of Population Exposure to Assess Clinical Safety (March 1995)*
E2A *Clinical Safety Data Management: Definitions and Standards for Expedited Reporting (March 1995)*
E2B *Clinical Safety Data Management: Data Elements for Transmission of Individual Case Safety Reports (January 1998)*
E2C *Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs (November 1996)*
E3 *Structure and Content of Clinical Study Reports (July 1996)*
E4 *Dose-Response Information to Support Drug Registration (November 1994)*
E5 *Ethnic Factors in the Acceptability of Foreign Clinical Data (June 1998)*
E6 *Good Clinical Practice: Consolidated Guideline (April 1996)*
E7 *Studies in Support of Special Populations: Geriatrics (August 1994)*
E8 *General Considerations for Clinical Trials (December 1997*
E10 *Choice of Control Group in Clinical Trials (September 1999)*
M1 *Standardization of Medical Terminology for Regulatory Purposes (November 1999)*
M3 *Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals (July 1997)*

This guidance is intended to give direction to sponsors in the design, conduct, analysis, and evaluation of clinical trials of an investigational product in the context of its overall clinical development. The document will also assist scientific experts charged with preparing application summaries or assessing evidence of efficacy and safety, principally from clinical trials in later phases of development.

B. Scope and Direction (1.2)

The focus of this guidance is on statistical principles. It does not address the use of specific statistical procedures or methods. Specific procedural steps to ensure that principles are implemented properly are the responsibility of the sponsor. Integration of data across clinical trials is discussed, but is not a primary focus of this guidance. Selected principles and procedures related to data management or clinical trial monitoring activities are covered in other ICH guidances and are not addressed here.

This guidance should be of interest to individuals from a broad range of scientific disciplines. However, it is assumed that the actual responsibility for all statistical work associated with clinical trials will lie with an appropriately qualified and experienced statistician, as indicated in ICH E6. The role and responsibility of the trial statistician (see Glossary), in collaboration with other clinical trial professionals, is to ensure that statistical principles are applied appropriately in clinical trials supporting drug

development. Thus, the trial statistician should have a combination of education/training and experience sufficient to implement the principles articulated in this guidance.

For each clinical trial contributing to a marketing application, all important details of its design and conduct and the principal features of its proposed statistical analysis should be clearly specified in a protocol written before the trial begins. The extent to which the procedures in the protocol are followed and the primary analysis is planned a priori will contribute to the degree of confidence in the final results and conclusions of the trial. The protocol and subsequent amendments should be approved by the responsible personnel, including the trial statistician. The trial statistician should ensure that the protocol and any amendments cover all relevant statistical issues clearly and accurately, using technical terminology as appropriate.

The principles outlined in this guidance are primarily relevant to clinical trials conducted in the later phases of development, many of which are confirmatory trials of efficacy. In addition to efficacy, confirmatory trials may have as their primary variable a safety variable (e.g., an adverse event, a clinical laboratory variable, or an electrocardiographic measure) or a pharmacodynamic or pharmacokinetic variable (as in a confirmatory bioequivalence trial). Furthermore, some confirmatory findings may be derived from data integrated across trials, and selected principles in this guidance are applicable in this situation. Finally, although the early phases of drug development consist mainly of clinical trials that are exploratory in nature, statistical principles are also relevant to these clinical trials. Hence, the substance of this document should be applied as far as possible to all phases of clinical development.

Many of the principles delineated in this guidance deal with minimizing bias (see Glossary) and maximizing precision. As used in this guidance, the term *bias* describes the systematic tendency of any factors associated with the design, conduct, analysis, and interpretation of the results of clinical trials to make the estimate of a treatment effect (see Glossary) deviate from its true value. It is important to identify potential sources of bias as completely as possible so that attempts to limit such bias may be made. The presence of bias may seriously compromise the ability to draw valid conclusions from clinical trials.

Some sources of bias arise from the design of the trial, for example an assignment of treatments such that subjects at lower risk are systematically assigned to one treatment. Other sources of bias arise during the conduct and analysis of a clinical trial. For example, protocol violations and exclusion of subjects from analysis based upon knowledge of subject outcomes are possible sources of bias that may affect the accurate assessment of the treatment effect. Because bias can occur in subtle or unknown ways and its effect is not measurable directly, it is important to evaluate the robustness of the results and primary conclusions of the trial. Robustness is a concept that refers to the sensitivity of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis. Robustness implies that the treatment effect and primary conclusions of the trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches. The interpretation of statistical measures of

uncertainty of the treatment effect and treatment comparisons should involve consideration of the potential contribution of bias to the p-value, confidence interval, or inference.

Because the predominant approaches to the design and analysis of clinical trials have been based on frequentist statistical methods, the guidance largely refers to the use of frequentist methods (see Glossary) when discussing hypothesis testing and/or confidence intervals. This should not be taken to imply that other approaches are not appropriate; the use of Bayesian (see Glossary) and other approaches may be considered when the reasons for their use are clear and when the resulting conclusions are sufficiently robust.

II. CONSIDERATIONS FOR OVERALL CLINICAL DEVELOPMENT

A. Trial Context (2.1)

1. Development Plan (2.1.1)

The broad aim of the process of clinical development of a new drug is to find out whether there is a dose range and schedule at which the drug can be shown to be simultaneously safe and effective, to the extent that the risk-benefit relationship is acceptable. The particular subjects who may benefit from the drug, and the specific indications for its use, also need to be defined.

Satisfying these broad aims usually requires an ordered program of clinical trials, each with its own specific objectives (see ICH E8). This should be specified in a clinical plan, or a series of plans, with appropriate decision points and flexibility to allow modification as knowledge accumulates. A marketing application should clearly describe the main content of such plans, and the contribution made by each trial. Interpretation and assessment of the evidence from the total program of trials involves synthesis of the evidence from the individual trials (see section VII.B). This is facilitated by ensuring that common standards are adopted for a number of features of the trials, such as dictionaries of medical terms, definition and timing of the main measurements, handling of protocol deviations, and so on. A statistical summary, overview, or meta-analysis (see Glossary) may be informative when medical questions are addressed in more than one trial. Where possible, this should be envisaged in the plan so that the relevant trials are clearly identified and any necessary common features of their designs are specified in advance. Other major statistical issues (if any) that are expected to affect a number of trials in a common plan should be addressed in that plan.

2. Confirmatory Trial (2.1.2)

A confirmatory trial is an adequately controlled trial in which the hypotheses are stated in advance and evaluated. As a rule, confirmatory trials are necessary to provide firm evidence of efficacy or safety. In such trials the key hypothesis of interest follows directly from the trial's primary objective, is always predefined,

and is the hypothesis that is subsequently tested when the trial is complete. In a confirmatory trial, it is equally important to estimate with due precision the size of the effects attributable to the treatment of interest and to relate these effects to their clinical significance.

Confirmatory trials are intended to provide firm evidence in support of claims; hence adherence to protocols and standard operating procedures is particularly important. Unavoidable changes should be explained and documented, and their effect examined. A justification of the design of each such trial and of other important statistical aspects, such as the principal features of the planned analysis, should be set out in the protocol. Each trial should address only a limited number of questions.

Firm evidence in support of claims requires that the results of the confirmatory trials demonstrate that the investigational product under test has clinical benefits. The confirmatory trials should therefore be sufficient to answer each key clinical question relevant to the efficacy or safety claim clearly and definitively. In addition, it is important that the basis for generalization (see Glossary) to the intended patient population is understood and explained; this may also influence the number and type (e.g., specialist or general practitioner) of centers and/or trials needed. The results of the confirmatory trial(s) should be robust. In some circumstances, the weight of evidence from a single confirmatory trial may be sufficient.

3. Exploratory Trial (2.1.3)

The rationale and design of confirmatory trials nearly always rests on earlier clinical work carried out in a series of exploratory studies. Like all clinical trials, these exploratory studies should have clear and precise objectives. However, in contrast to confirmatory trials, their objectives may not always lead to simple tests of predefined hypotheses. In addition, exploratory trials may sometimes require a more flexible approach to design so that changes can be made in response to accumulating results. Their analysis may entail data exploration. Tests of hypothesis may be carried out, but the choice of hypothesis may be data dependent. Such trials cannot be the basis of the formal proof of efficacy, although they may contribute to the total body of relevant evidence.

Any individual trial may have both confirmatory and exploratory aspects. For example, in most confirmatory trials the data are also subjected to exploratory analyses which serve as a basis for explaining or supporting their findings and for suggesting further hypotheses for later research. The protocol should make a clear distinction between the aspects of a trial which will be used for confirmatory proof and the aspects which will provide data for exploratory analysis.

B. Scope of Trials (2.2)

1. Population (2.2.1)

In the earlier phases of drug development, the choice of subjects for a clinical trial may be heavily influenced by the wish to maximize the chance of observing specific clinical effects of interest. Hence they may come from a very narrow subgroup of the total patient population for which the drug may eventually be indicated. However, by the time the confirmatory trials are undertaken, the subjects in the trials should more closely mirror the target population. In these trials, it is generally helpful to relax the inclusion and exclusion criteria as much as possible within the target population while maintaining sufficient homogeneity to permit precise estimation of treatment effects. No individual clinical trial can be expected to be totally representative of future users because of the possible influences of geographical location, the time when it is conducted, the medical practices of the particular investigator(s) and clinics, and so on. However, the influence of such factors should be reduced wherever possible and subsequently discussed during the interpretation of the trial results.

2. Primary and Secondary Variables (2.2.2)

The primary variable (*target* variable, primary endpoint) should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial. There should generally be only one primary variable. This will usually be an efficacy variable, because the primary objective of most confirmatory trials is to provide strong scientific evidence regarding efficacy. Safety/tolerability may sometimes be the primary variable, and will always be an important consideration. Measurements relating to quality of life and health economics are further potential primary variables. The selection of the primary variable should reflect the accepted norms and standards in the relevant field of research. The use of a reliable and validated variable with which experience has been gained either in earlier studies or in published literature is recommended. There should be sufficient evidence that the primary variable can provide a valid and reliable measure of some clinically relevant and important treatment benefit in the patient population described by the inclusion and exclusion criteria. The primary variable should generally be the one used when estimating the sample size (see section III.E).

In many cases, the approach to assessing subject outcome may not be straightforward and should be carefully defined. For example, it is inadequate to specify mortality as a primary variable without further clarification; mortality may be assessed by comparing proportions alive at fixed points in time or by comparing overall distributions of survival times over a specified interval. Another common example is a recurring event; the measure of treatment effect may again be a simple dichotomous variable (any occurrence during a specified interval), time to first

occurrence, rate of occurrence (events per time units of observation), and so on. The assessment of functional status over time in studying treatment for chronic disease presents other challenges in selection of the primary variable. There are many possible approaches, such as comparisons of the assessments done at the beginning and end of the interval of observation, comparisons of slopes calculated from all assessments throughout the interval, comparisons of the proportions of subjects exceeding or declining beyond a specified threshold, or comparisons based on methods for repeated measures data. To avoid multiplicity concerns arising from post hoc definitions, it is critical to specify in the protocol the precise definition of the primary variable as it will be used in the statistical analysis. In addition, the clinical relevance of the specific primary variable selected and the validity of the associated measurement procedures will generally need to be addressed and justified in the protocol.

The primary variable should be specified in the protocol, along with the rationale for its selection. Redefinition of the primary variable after unblinding will almost always be unacceptable, since the biases this introduces are difficult to assess. When the clinical effect defined by the primary objective is to be measured in more than one way, the protocol should identify one of the measurements as the primary variable on the basis of clinical relevance, importance, objectivity, and/or other relevant characteristics, whenever such selection is feasible.

Secondary variables are either supportive measurements related to the primary objective or measurements of effects related to the secondary objectives. Their predefinition in the protocol is also important, as well as an explanation of their relative importance and roles in interpretation of trial results. The number of secondary variables should be limited and should be related to the limited number of questions to be answered in the trial.

3. *Composite Variables (2.2.3)*

If a single primary variable cannot be selected from multiple measurements associated with the primary objective, another useful strategy is to integrate or combine the multiple measurements into a single or *composite* variable, using a predefined algorithm. Indeed, the primary variable sometimes arises as a combination of multiple clinical measurements (e.g., the rating scales used in arthritis, psychiatric disorders, and elsewhere). This approach addresses the multiplicity problem without requiring adjustment to the Type I error. The method of combining the multiple measurements should be specified in the protocol, and an interpretation of the resulting scale should be provided in terms of the size of a clinically relevant benefit. When a composite variable is used as a primary variable, the components of this variable may sometimes be analyzed separately, where clinically meaningful and validated. When a rating scale is used as a primary variable, it is especially important to address factors such as content

validity (see Glossary), inter- and intrarater reliability (see Glossary), and responsiveness for detecting changes in the severity of disease.

4. Global Assessment Variables (2.2.4)

In some cases, *global assessment* variables (see Glossary) are developed to measure the overall safety, overall efficacy, and/or overall usefulness of a treatment. This type of variable integrates objective variables and the investigator's overall impression about the state or change in the state of the subject, and is usually a scale of ordered categorical ratings. Global assessments of overall efficacy are well established in some therapeutic areas, such as neurology and psychiatry.

Global assessment variables generally have a subjective component. When a global assessment variable is used as a primary or secondary variable, fuller details of the scale should be included in the protocol with respect to:

- The relevance of the scale to the primary objective of the trial;
- The basis for the validity and reliability of the scale;
- How to utilize the data collected on an individual subject to assign him/her to a unique category of the scale;
- How to assign subjects with missing data to a unique category of the scale, or otherwise evaluate them.

If objective variables are considered by the investigator when making a global assessment, then those objective variables should be considered as additional primary or, at least, important secondary variables.

Global assessment of usefulness integrates components of both benefit and risk and reflects the decisionmaking process of the treating physician, who must weigh benefit and risk in making product use decisions. A problem with global usefulness variables is that their use could in some cases lead to the result of two products being declared equivalent despite having very different profiles of beneficial and adverse effects. For example, judging the global usefulness of a treatment as equivalent or superior to an alternative may mask the fact that it has little or no efficacy but fewer adverse effects. Therefore, it is not advisable to use a global usefulness variable as a primary variable. If global usefulness is specified as primary, it is important to consider specific efficacy and safety outcomes separately as additional primary variables.

5. *Multiple Primary Variables (2.2.5)*

It may sometimes be desirable to use more than one primary variable, each of which (or a subset of which) could be sufficient to cover the range of effects of the therapies. The planned manner of interpretation of this type of evidence should be carefully spelled out. It should be clear whether an impact on any of the variables, some minimum number of them, or all of them, would be considered necessary to achieve the trial objectives. The primary hypothesis or hypotheses and parameters of interest (e.g., mean, percentage, distribution) should be clearly stated with respect to the primary variables identified, and the approach to statistical inference described. The effect on the Type I error should be explained because of the potential for multiplicity problems (see section V.F); the method of controlling Type I error should be given in the protocol. The extent of intercorrelation among the proposed primary variables may be considered in evaluating the impact on Type I error. If the purpose of the trial is to demonstrate effects on all of the designated primary variables, then there is no need for adjustment of the Type I error, but the impact on Type II error and sample size should be carefully considered.

6. *Surrogate Variables (2.2.6)*

When direct assessment of the clinical benefit to the subject through observing actual clinical efficacy is not practical, indirect criteria (surrogate variables — see Glossary) may be considered. Commonly accepted surrogate variables are used in a number of indications where they are believed to be reliable predictors of clinical benefit. There are two principal concerns with the introduction of any proposed surrogate variable. First, it may not be a true predictor of the clinical outcome of interest. For example, it may measure treatment activity associated with one specific pharmacological mechanism, but may not provide full information on the range of actions and ultimate effects of the treatment, whether positive or negative. There have been many instances where treatments showing a highly positive effect on a proposed surrogate have ultimately been shown to be detrimental to the subjects' clinical outcome; conversely, there are cases of treatments conferring clinical benefit without measurable impact on proposed surrogates. Second, proposed surrogate variables may not yield a quantitative measure of clinical benefit that can be weighed directly against adverse effects. Statistical criteria for validating surrogate variables have been proposed but the experience with their use is relatively limited. In practice, the strength of the evidence for surrogacy depends upon (i) the biological plausibility of the relationship, (ii) the demonstration in epidemiological studies of the prognostic value of the surrogate for the clinical outcome, and (iii) evidence from clinical trials that treatment effects on the surrogate correspond to effects on the clinical outcome. Relationships between clinical and surrogate variables for one product do not necessarily apply to a product with a different mode of action for treating the same disease.

7. *Categorized Variables (2.2.7)*

Dichotomization or other categorization of continuous or ordinal variables may sometimes be desirable. Criteria of *success* and *response* are common examples of dichotomies that should be specified precisely in terms of, for example, a minimum percentage improvement (relative to baseline) in a continuous variable or a ranking categorized as at or above some threshold level (e.g., *good*) on an ordinal rating scale. The reduction of diastolic blood pressure below 90 mmHg is a common dichotomization. Categorizations are most useful when they have clear clinical relevance. The criteria for categorization should be predefined and specified in the protocol, as knowledge of trial results could easily bias the choice of such criteria. Because categorization normally implies a loss of information, a consequence will be a loss of power in the analysis; this should be accounted for in the sample size calculation.

C. Design Techniques to Avoid Bias (2.3)

The most important design techniques for avoiding bias in clinical trials are blinding and randomization, and these should be normal features of most controlled clinical trials intended to be included in a marketing application. Most such trials follow a double-blind approach in which treatments are prepacked in accordance with a suitable randomization schedule, and supplied to the trial center(s) labeled only with the subject number and the treatment period, so that no one involved in the conduct of the trial is aware of the specific treatment allocated to any particular subject, not even as a code letter. This approach will be assumed in section II.C.1 and most of section II.C.2, exceptions being considered at the end.

Bias can also be reduced at the design stage by specifying procedures in the protocol aimed at minimizing any anticipated irregularities in trial conduct that might impair a satisfactory analysis, including various types of protocol violations, withdrawals and missing values. The protocol should consider ways both to reduce the frequency of such problems and to handle the problems that do occur in the analysis of data.

1. *Blinding (2.3.1)*

Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence that the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of end-points, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.

A double-blind trial is one in which neither the subject nor any of the investigator or sponsor staff involved in the treatment or clinical evaluation of the subjects are aware of the treatment received. This includes anyone determining subject eligibility, evaluating endpoints, or assessing compliance with the protocol. This level of blinding is maintained throughout the conduct of the trial, and only when the data are cleaned to an acceptable level of quality will appropriate personnel be unblinded. If any of the sponsor staff who are not involved in the treatment or clinical evaluation of the subjects are required to be unblinded to the treatment code (e.g., bioanalytical scientists, auditors, those involved in serious adverse event reporting), the sponsor should have adequate standard operating procedures to guard against inappropriate dissemination of treatment codes. In a single-blind trial the investigator and/or his staff are aware of the treatment but the subject is not, or vice versa. In an open-label trial the identity of treatment is known to all. The double-blind trial is the optimal approach. This requires that the treatments to be applied during the trial cannot be distinguished (by appearance, taste, etc.) either before or during administration, and that the blind is maintained appropriately during the whole trial.

Difficulties in achieving the double-blind ideal can arise: The treatments may be of a completely different nature, for example, surgery and drug therapy; two drugs may have different formulations and, although they could be made indistinguishable by the use of capsules, changing the formulation might also change the pharmacokinetic and/or pharmacodynamic properties and hence necessitate that bioequivalence of the formulations be established; the daily pattern of administration of two treatments may differ. One way of achieving double-blind conditions under these circumstances is to use a *double-dummy* (see Glossary) technique. This technique may sometimes force an administration scheme that is sufficiently unusual to influence adversely the motivation and compliance of the subjects. Ethical difficulties may also interfere with its use when, for example, it entails dummy operative procedures. Nevertheless, extensive efforts should be made to overcome these difficulties.

The double-blind nature of some clinical trials may be partially compromised by apparent treatment induced effects. In such cases, blinding may be improved by blinding investigators and relevant sponsor staff to certain test results (e.g., selected clinical laboratory measures). Similar approaches (see below) to minimizing bias in open-label trials should be considered in trials where unique or specific treatment effects may lead to unblinding individual patients.

If a double-blind trial is not feasible, then the single-blind option should be considered. In some cases only an open-label trial is practically or ethically possible. Single-blind and open-label trials provide additional flexibility, but it is particularly important that the investigator's knowledge of the next treatment should not influence the decision to enter the subject; this decision should precede knowledge of the randomized treatment. For these trials, consideration should be given to the use of a centralized randomization method, such as telephone

randomization, to administer the assignment of randomized treatment. In addition, clinical assessments should be made by medical staff who are not involved in treating the subjects and who remain blind to treatment. In single-blind or open-label trials every effort should be made to minimize the various known sources of bias and primary variables should be as objective as possible. The reasons for the degree of blinding adopted, as well as steps taken to minimize bias by other means, should be explained in the protocol. For example, the sponsor should have adequate standard operating procedures to ensure that access to the treatment code is appropriately restricted during the process of cleaning the database prior to its release for analysis.

Breaking the blind (for a single subject) should be considered only when knowledge of the treatment assignment is deemed essential by the subject's physician for the subject's care. Any intentional or unintentional breaking of the blind should be reported and explained at the end of the trial, irrespective of the reason for its occurrence. The procedure and timing for revealing the treatment assignments should be documented.

In this document, the blind review (see Glossary) of data refers to the checking of data during the period of time between trial completion (the last observation on the last subject) and the breaking of the blind.

2. Randomization (2.3.2)

Randomization introduces a deliberate element of chance into the assignment of treatments to subjects in a clinical trial. During subsequent analysis of the trial data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. It also tends to produce treatment groups in which the distributions of prognostic factors, known and unknown, are similar. In combination with blinding, randomization helps to avoid possible bias in the selection and allocation of subjects arising from the predictability of treatment assignments.

The randomization schedule of a clinical trial documents the random allocation of treatments to subjects. In the simplest situation it is a sequential list of treatments (or treatment sequences in a crossover trial) or corresponding codes by subject number. The logistics of some trials, such as those with a screening phase, may make matters more complicated, but the unique preplanned assignment of treatment, or treatment sequence, to subject should be clear. Different trial designs will necessitate different procedures for generating randomization schedules. The randomization schedule should be reproducible (if the need arises).

Although unrestricted randomization is an acceptable approach, some advantages can generally be gained by randomizing subjects in blocks. This helps to increase the comparability of the treatment groups, particularly when subject characteristics may change over time, as a result, for example, of changes in recruitment policy. It also provides a better guarantee that the treatment groups will be of nearly equal

size. In crossover trials, it provides the means of obtaining balanced designs with their greater efficiency and easier interpretation. Care should be taken to choose block lengths that are sufficiently short to limit possible imbalance, but that are long enough to avoid predictability towards the end of the sequence in a block. Investigators and other relevant staff should generally be blind to the block length; the use of two or more block lengths, randomly selected for each block, can achieve the same purpose. (Theoretically, in a double-blind trial predictability does not matter, but the pharmacological effects of drugs may provide the opportunity for intelligent guesswork.)

In multicenter trials (see Glossary), the randomization procedures should be organized centrally. It is advisable to have a separate random scheme for each center, i.e., to stratify by center or to allocate several whole blocks to each center. More generally, stratification by important prognostic factors measured at baseline (e.g., severity of disease, age, sex) may sometimes be valuable in order to promote balanced allocation within strata; this has greater potential benefit in small trials. The use of more than two or three stratification factors is rarely necessary, is less successful at achieving balance, and is logistically troublesome. The use of a dynamic allocation procedure (see below) may help to achieve balance across a number of stratification factors simultaneously, provided the rest of the trial procedures can be adjusted to accommodate an approach of this type. Factors on which randomization has been stratified should be accounted for later in the analysis.

The next subject to be randomized into a trial should always receive the treatment corresponding to the next free number in the appropriate randomization schedule (in the respective stratum, if randomization is stratified). The appropriate number and associated treatment for the next subject should only be allocated when entry of that subject to the randomized part of the trial has been confirmed. Details of the randomization that facilitate predictability (e.g., block length) should not be contained in the trial protocol. The randomization schedule itself should be filed securely by the sponsor or an independent party in a manner that ensures that blindness is properly maintained throughout the trial. Access to the randomization schedule during the trial should take into account the possibility that, in an emergency, the blind may have to be broken for any subject. The procedure to be followed, the necessary documentation, and the subsequent treatment and assessment of the subject should all be described in the protocol.

Dynamic allocation is an alternative procedure in which the allocation of treatment to a subject is influenced by the current balance of allocated treatments and, in a stratified trial, by the stratum to which the subject belongs and the balance within that stratum. Deterministic dynamic allocation procedures should be avoided and an appropriate element of randomization should be incorporated for each treatment allocation. Every effort should be made to retain the double-blind status of the trial. For example, knowledge of the treatment code may be restricted to a central trial office from where the dynamic allocation is controlled, generally through telephone

contact. This in turn permits additional checks of eligibility criteria and establishes entry into the trial, features that can be valuable in certain types of multicenter trials. The usual system of prepacking and labeling drug supplies for double-blind trials can then be followed, but the order of their use is no longer sequential. It is desirable to use appropriate computer algorithms to keep personnel at the central trial office blind to the treatment code. The complexity of the logistics and potential impact on the analysis should be carefully evaluated when considering dynamic allocation.

III. TRIAL DESIGN CONSIDERATIONS

A. Design Configuration (3.1)

1. Parallel Group Design (3.1.1)

The most common clinical trial design for confirmatory trials is the parallel group design in which subjects are randomized to one of two or more arms, each arm being allocated a different treatment. These treatments will include the investigational product at one or more doses, and one or more control treatments, such as placebo and/or an active comparator. The assumptions underlying this design are less complex than for most other designs. However, as with other designs, there may be additional features of the trial that complicate the analysis and interpretation (e.g., covariates, repeated measurements over time, interactions between design factors, protocol violations, dropouts (see Glossary), and withdrawals).

2. Crossover Design (3.1.2)

In the crossover design, each subject is randomized to a sequence of two or more treatments and hence acts as his own control for treatment comparisons. This simple maneuver is attractive primarily because it reduces the number of subjects and usually the number of assessments needed to achieve a specific power, sometimes to a marked extent. In the simplest 2x2 crossover design, each subject receives each of two treatments in randomized order in two successive treatment periods, often separated by a washout period. The most common extension of this entails comparing n(>2) treatments in n periods, each subject receiving all n treatments. Numerous variations exist, such as designs in which each subject receives a subset of n(>2) treatments, or designs in which treatments are repeated within a subject.

Crossover designs have a number of problems that can invalidate their results. The chief difficulty concerns carryover, that is, the residual influence of treatments in subsequent treatment periods. In an additive model, the effect of unequal carryover will be to bias direct treatment comparisons. In the 2x2 design, the carryover effect cannot be statistically distinguished from the interaction between treatment and

period and the test for either of these effects lacks power because the corresponding contrast is *between subject*. This problem is less acute in higher order designs, but cannot be entirely dismissed.

When the crossover design is used, it is therefore important to avoid carryover. This is best done by selective and careful use of the design on the basis of adequate knowledge of both the disease area and the new medication. The disease under study should be chronic and stable. The relevant effects of the medication should develop fully within the treatment period. The washout periods should be sufficiently long for complete reversibility of drug effect. The fact that these conditions are likely to be met should be established in advance of the trial by means of prior information and data.

There are additional problems that need careful attention in crossover trials. The most notable of these are the complications of analysis and interpretation arising from the loss of subjects. Also, the potential for carryover leads to difficulties in assigning adverse events that occur in later treatment periods to the appropriate treatment. These and other issues are described in ICH E4. The crossover design should generally be restricted to situations where losses of subjects from the trial are expected to be small.

A common, and generally satisfactory, use of the 2x2 crossover design is to demonstrate the bioequivalence of two formulations of the same medication. In this particular application in healthy volunteers, carryover effects on the relevant pharmacokinetic variable are most unlikely to occur if the wash-out time between the two periods is sufficiently long. However, it is still important to check this assumption during analysis on the basis of the data obtained, for example, by demonstrating that no drug is detectable at the start of each period.

3. *Factorial Designs (3.1.3)*

In a factorial design, two or more treatments are evaluated simultaneously through the use of varying combinations of the treatments. The simplest example is the 2x2 factorial design in which subjects are randomly allocated to one of the four possible combinations of two treatments, A and B. These are: A alone; B alone; both A and B; neither A nor B. In many cases, this design is used for the specific purpose of examining the interaction of A and B. The statistical test of interaction may lack power to detect an interaction if the sample size was calculated based on the test for main effects. This consideration is important when this design is used for examining the joint effects of A and B, in particular, if the treatments are likely to be used together.

Another important use of the factorial design is to establish the dose-response characteristics of the simultaneous use of treatments C and D, especially when the efficacy of each monotherapy has been established at some dose in prior trials. A number, m, of doses of C is selected, usually including a zero dose (placebo), and a

similar number, n, of doses of D. The full design then consists of mxn treatment groups, each receiving a different combination of doses of C and D. The resulting estimate of the response surface may then be used to help identify an appropriate combination of doses of C and D for clinical use (see ICH E4).

In some cases, the 2x2 design may be used to make efficient use of clinical trial subjects by evaluating the efficacy of the two treatments with the same number of subjects as would be required to evaluate the efficacy of either one alone. This strategy has proved to be particularly valuable for very large mortality trials. The efficiency and validity of this approach depends upon the absence of interaction between treatments A and B so that the effects of A and B on the primary efficacy variables follow an additive model. Hence the effect of A is virtually identical whether or not it is additional to the effect of B. As for the crossover trial, evidence that this condition is likely to be met should be established in advance of the trial by means of prior information and data.

B. Multicenter Trials (3.2)

Multicenter trials are carried out for two main reasons. First, a multicenter trial is an accepted way of evaluating a new medication more efficiently. Under some circumstances, it may present the only practical means of accruing sufficient subjects to satisfy the trial objective within a reasonable timeframe. Multicenter trials of this nature may, in principle, be carried out at any stage of clinical development. They may have several centers with a large number of subjects per center or, in the case of a rare disease, they may have a large number of centers with very few subjects per center.

Second, a trial may be designed as a multicenter (and multi-investigator) trial primarily to provide a better basis for the subsequent generalization of its findings. This arises from the possibility of recruiting the subjects from a wider population and of administering the medication in a broader range of clinical settings, thus presenting an experimental situation that is more typical of future use. In this case, the involvement of a number of investigators also gives the potential for a wider range of clinical judgement concerning the value of the medication. Such a trial would be a confirmatory trial in the later phases of drug development and would be likely to involve a large number of investigators and centers. It might sometimes be conducted in a number of different countries to facilitate generalizability (see Glossary) even further.

If a multicenter trial is to be meaningfully interpreted and extrapolated, then the manner in which the protocol is implemented should be clear and similar at all centers. Furthermore, the usual sample size and power calculations depend upon the assumption that the differences between the compared treatments in the centers are unbiased estimates of the same quantity. It is important to design the common protocol and to conduct the trial with this background in mind. Procedures should be standardized as completely as possible. Variation of evaluation criteria and schemes can be reduced by investigator meetings, by the training of personnel in advance of the trial, and by careful monitoring during the trial.

Good design should generally aim to achieve the same distribution of subjects to treatments within each center and good management should maintain this design objective. Trials that avoid excessive variation in the numbers of subjects per center and trials that avoid a few very small centers have advantages if it is later found necessary to take into account the heterogeneity of the treatment effect from center to center, because they reduce the differences between different weighted estimates of the treatment effect. (This point does not apply to trials in which all centers are very small and in which center does not feature in the analysis.) Failure to take these precautions, combined with doubts about the homogeneity of the results, may, in severe cases, reduce the value of a multicenter trial to such a degree that it cannot be regarded as giving convincing evidence for the sponsor's claims.

In the simplest multicenter trial, each investigator will be responsible for the subjects recruited at one hospital, so that *center* is identified uniquely by either investigator or hospital. In many trials, however, the situation is more complex. One investigator may recruit subjects from several hospitals; one investigator may represent a team of clinicians (subinvestigators) who all recruit subjects from their own clinics at one hospital or at several associated hospitals. Whenever there is room for doubt about the definition of center in a statistical model, the statistical section of the protocol (see section V.A) should clearly define the term (e.g., by investigator, location or region) in the context of the particular trial. In most instances, centers can be satisfactorily defined through the investigators. (ICH E6 provides relevant guidance in this respect.) In cases of doubt, the aim should be to define centers to achieve homogeneity in the important factors affecting the measurements of the primary variables and the influence of the treatments. Any rules for combining centers in the analysis should be justified and specified prospectively in the protocol where possible, but in any case decisions concerning this approach should always be taken blind to treatment, for example, at the time of the blind review.

The statistical model to be adopted for the estimation and testing of treatment effects should be described in the protocol. The main treatment effect may be investigated first using a model that allows for center differences, but does not include a term for treatment-by-center interaction. If the treatment effect is homogeneous across centers, the routine inclusion of interaction terms in the model reduces the efficiency of the test for the main effects. In the presence of true heterogeneity of treatment effects, the interpretation of the main treatment effect is controversial.

In some trials, for example, some large mortality trials with very few subjects per center, there may be no reason to expect the centers to have any influence on the primary or secondary variables because they are unlikely to represent influences of clinical importance. In other trials, it may be recognized from the start that the limited numbers of subjects per center will make it impracticable to include the center effects in the statistical model. In these cases, it is not considered appropriate to include a term for center in the model, and it is not necessary to stratify the randomization by center in this situation.

If positive treatment effects are found in a trial with appreciable numbers of subjects per center, there should generally be an exploration of the heterogeneity of treatment effects

across centers, as this may affect the generalizability of the conclusions. Marked heterogeneity may be identified by graphical display of the results of individual centers or by analytical methods, such as a significance test of the treatment-by-center interaction. When using such a statistical significance test, it is important to recognize that this generally has low power in a trial designed to detect the main effect of treatment.

If heterogeneity of treatment effects is found, this should be interpreted with care, and vigorous attempts should be made to find an explanation in terms of other features of trial management or subject characteristics. Such an explanation will usually suggest appropriate further analysis and interpretation. In the absence of an explanation, heterogeneity of treatment effect, as evidenced, for example, by marked quantitative interactions (see Glossary) implies that alternative estimates of the treatment effect, giving different weights to the centers, may be needed to substantiate the robustness of the estimates of treatment effect. It is even more important to understand the basis of any heterogeneity characterized by marked qualitative interactions (see Glossary), and failure to find an explanation may necessitate further clinical trials before the treatment effect can be reliably predicted.

Up to this point, the discussion of multicenter trials has been based on the use of fixed effect models. Mixed models may also be used to explore the heterogeneity of the treatment effect. These models consider center and treatment-by-center effects to be random and are especially relevant when the number of sites is large.

C. Type of Comparison (3.3)

1. Trials to Show Superiority (3.3.1)

Scientifically, efficacy is most convincingly established by demonstrating superiority to placebo in a placebo-controlled trial, by showing superiority to an active control treatment, or by demonstrating a dose-response relationship. This type of trial is referred to as a *superiority* trial (see Glossary). In this guidance superiority trials are generally assumed, unless explicitly stated otherwise.

For serious illnesses, when a therapeutic treatment that has been shown to be efficacious by superiority trial(s) exists, a placebo-controlled trial may be considered unethical. In that case the scientifically sound use of an active treatment as a control should be considered. The appropriateness of placebo control versus active control should be considered on a trial-by-trial basis.

2. Trials to Show Equivalence or Noninferiority (3.3.2)

In some cases, an investigational product is compared to a reference treatment without the objective of showing superiority. This type of trial is divided into two major categories according to its objective; one is an *equivalence* trial (see Glossary) and the other is a *noninferiority* trial (see Glossary).

Bioequivalence trials fall into the former category. In some situations, clinical equivalence trials are also undertaken for other regulatory reasons such as demonstrating the clinical equivalence of a generic product to the marketed product when the compound is not absorbed and therefore not present in the blood stream.

Many active control trials are designed to show that the efficacy of an investigational product is no worse than that of the active comparator and, hence, fall into the latter category. Another possibility is a trial in which multiple doses of the investigational drug are compared with the recommended dose or multiple doses of the standard drug. The purpose of this design is simultaneously to show a dose-response relationship for the investigational product and to compare the investigational product with the active control.

Active control equivalence or noninferiority trials may also incorporate a placebo, thus pursuing multiple goals in one trial. For example, they may establish superiority to placebo and hence validate the trial design and simultaneously evaluate the degree of similarity of efficacy and safety to the active comparator. There are well-known difficulties associated with the use of the active control equivalence (or noninferiority) trials that do not incorporate a placebo or do not use multiple doses of the new drug. These relate to the implicit lack of any measure of internal validity (in contrast to superiority trials), thus making external validation necessary. The equivalence (or noninferiority) trial is not conservative in nature, so that many flaws in the design or conduct of the trial will tend to bias the results towards a conclusion of equivalence. For these reasons, the design features of such trials should receive special attention and their conduct needs special care. For example, it is especially important to minimize the incidence of violations of the entry criteria, noncompliance, withdrawals, losses to follow-up, missing data, and other deviations from the protocol, and also to minimize their impact on the subsequent analyses.

Active comparators should be chosen with care. An example of a suitable active comparator would be a widely used therapy whose efficacy in the relevant indication has been clearly established and quantified in well-designed and well-documented superiority trial(s) and that can be reliably expected to exhibit similar efficacy in the contemplated active control trial. To this end, the new trial should have the same important design features (primary variables, the dose of the active comparator, eligibility criteria, and so on) as the previously conducted superiority trials in which the active comparator clearly demonstrated clinically relevant efficacy, taking into account advances in medical or statistical practice relevant to the new trial.

It is vital that the protocol of a trial designed to demonstrate equivalence or noninferiority contain a clear statement that this is its explicit intention. An equivalence margin should be specified in the protocol; this margin is the largest difference that can be judged as being clinically acceptable and should be smaller

than differences observed in superiority trials of the active comparator. For the active control equivalence trial, both the upper and the lower equivalence margins are needed, while only the lower margin is needed for the active control noninferiority trial. The choice of equivalence margins should be justified clinically.

Statistical analysis is generally based on the use of confidence intervals (see section V.E). For equivalence trials, two-sided confidence intervals should be used. Equivalence is inferred when the entire confidence interval falls within the equivalence margins. Operationally, this is equivalent to the method of using two simultaneous one-sided tests to test the (composite) null hypothesis that the treatment difference is outside the equivalence margins versus the (composite) alternative hypothesis that the treatment difference is within the margins. Because the two null hypotheses are disjoint, the Type I error is appropriately controlled. For noninferiority trials, a one-sided interval should be used. The confidence interval approach has a one-sided hypothesis test counterpart for testing the null hypothesis that the treatment difference (investigational product minus control) is equal to the lower equivalence margin versus the alternative that the treatment difference is greater than the lower equivalence margin. The choice of Type I error should be a consideration separate from the use of a one-sided or two-sided procedure. Sample size calculations should be based on these methods (see section III.E).

Concluding equivalence or noninferiority based on observing a nonsignificant test result of the null hypothesis that there is no difference between the investigational product and the active comparator is considered inappropriate.

There are also special issues in the choice of analysis sets. Subjects who withdraw or drop out of the treatment group or the comparator group will tend to have a lack of response; hence the results of using the full analysis set (see Glossary) may be biased toward demonstrating equivalence (see section V.B.3).

3. Trials to Show Dose-Response Relationship (3.3.3)

How response is related to the dose of a new investigational product is a question to which answers may be obtained in all phases of development and by a variety of approaches (see ICH E4). Dose-response trials may serve a number of objectives, among which the following are of particular importance: the confirmation of efficacy; the investigation of the shape and location of the dose-response curve; the estimation of an appropriate starting dose; the identification of optimal strategies for individual dose adjustments; the determination of a maximal dose beyond which additional benefit would be unlikely to occur. These objectives should be addressed using the data collected at a number of doses under investigation, including a placebo (zero dose) wherever appropriate. For this purpose, the application of procedures to estimate the relationship between dose and response,

including the construction of confidence intervals and the use of graphical methods, is as important as the use of statistical tests. The hypothesis tests that are used may need to be tailored to the natural ordering of doses or to particular questions regarding the shape of the dose-response curve (e.g., monotonicity). The details of the planned statistical procedures should be given in the protocol.

D. Group Sequential Designs (3.4)

Group sequential designs are used to facilitate the conduct of interim analysis (see section IV.E and Glossary). While group sequential designs are not the only acceptable types of designs permitting interim analysis, they are the most commonly applied because it is more practicable to assess grouped subject outcomes at periodic intervals during the trial than on a continuous basis as data from each subject become available. The statistical methods should be fully specified in advance of the availability of information on treatment outcomes and subject treatment assignments (i.e., blind breaking, see section IV.E). An independent data monitoring committee (IDMC) (see Glossary) may be used to review or to conduct the interim analysis of data arising from a group sequential design (see section IV.F). While the design has been most widely and successfully used in large, long-term trials of mortality or major nonfatal endpoints, its use is growing in other circumstances. In particular, it is recognized that safety must be monitored in all trials; therefore, the need for formal procedures to cover early stopping for safety reasons should always be considered.

E. Sample Size (3.5)

The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed. This number is usually determined by the primary objective of the trial. If the sample size is determined on some other basis, then this should be made clear and justified. For example, a trial sized on the basis of safety questions or requirements or important secondary objectives may need larger numbers of subjects than a trial sized on the basis of the primary efficacy question (see ICH E1A).

Using the usual method for determining the appropriate sample size, the following items should be specified: A primary variable; the test statistic; the null hypothesis; the alternative (*working*) hypothesis at the chosen dose(s) (embodying consideration of the treatment difference to be detected or rejected at the dose and in the subject population selected); the probability of erroneously rejecting the null hypothesis (the Type I error) and the probability of erroneously failing to reject the null hypothesis (the Type II error); as well as the approach to dealing with treatment withdrawals and protocol violations. In some instances, the event rate is of primary interest for evaluating power, and assumptions should be made to extrapolate from the required number of events to the eventual sample size for the trial.

The method by which the sample size is calculated should be given in the protocol, together with the estimates of any quantities used in the calculations (such as variances, mean values, response rates, event rates, difference to be detected). The basis of these

estimates should also be given. It is important to investigate the sensitivity of the sample size estimate to a variety of deviations from these assumptions and this may be facilitated by providing a range of sample sizes appropriate for a reasonable range of deviations from assumptions. In confirmatory trials, assumptions should normally be based on published data or on the results of earlier trials. The treatment difference to be detected may be based on a judgement concerning the minimal effect which has clinical relevance in the management of patients or on a judgement concerning the anticipated effect of the new treatment, where this is larger. Conventionally, the probability of Type I error is set at 5 percent or less or as dictated by any adjustments made necessary for multiplicity considerations; the precise choice may be influenced by the prior plausibility of the hypothesis under test and the desired impact of the results. The probability of Type II error is conventionally set at 10 percent to 20 percent. It is in the sponsor's interest to keep this figure as low as feasible, especially in the case of trials that are difficult or impossible to repeat. Alternative values to the conventional levels of Type I and Type II error may be acceptable or even preferable in some cases.

Sample size calculations should refer to the number of subjects required for the primary analysis. If this is the *full analysis set*, estimates of the effect size may need to be reduced compared to the per protocol set (see Glossary). This is to allow for the dilution of the treatment effect arising from the inclusion of data from patients who have withdrawn from treatment or whose compliance is poor. The assumptions about variability may also need to be revised.

The sample size of an equivalence trial or a noninferiority trial (see section III.C.2) should normally be based on the objective of obtaining a confidence interval for the treatment difference that shows that the treatments differ at most by a clinically acceptable difference. When the power of an equivalence trial is assessed at a true difference of zero, then the sample size necessary to achieve this power is underestimated if the true difference is not zero. When the power of a noninferiority trial is assessed at a zero difference, then the sample size needed to achieve that power will be underestimated if the effect of the investigational product is less than that of the active control. The choice of a *clinically acceptable* difference needs justification with respect to its meaning for future patients, and may be smaller than the "clinically relevant" difference referred to above in the context of superiority trials designed to establish that a difference exists.

The exact sample size in a group sequential trial cannot be fixed in advance because it depends upon the play of chance in combination with the chosen stopping guideline and the true treatment difference. The design of the stopping guideline should take into account the consequent distribution of the sample size, usually embodied in the expected and maximum sample sizes.

When event rates are lower than anticipated or variability is larger than expected, methods for sample size reestimation are available without unblinding data or making treatment comparisons (see section IV.D).

F. Data Capture and Processing (3.6)

The collection of data and transfer of data from the investigator to the sponsor can take place through a variety of media, including paper case record forms, remote site monitoring systems, medical computer systems, and electronic transfer. Whatever data capture instrument is used, the form and content of the information collected should be in full accordance with the protocol and should be established in advance of the conduct of the clinical trial. It should focus on the data necessary to implement the planned analysis, including the context information (such as timing assessments relative to dosing) necessary to confirm protocol compliance or identify important protocol deviations. *Missing values* should be distinguishable from the *value zero* or *characteristic absent.*

The process of data capture, through to database finalization, should be carried out in accordance with good clinical practice (GCP) (see ICH E6, section 5). Specifically, timely and reliable processes for recording data and rectifying errors and omissions are necessary to ensure delivery of a quality database and the achievement of the trial objectives through the implementation of the planned analysis.

IV. TRIAL CONDUCT CONSIDERATIONS

A. Trial Monitoring and Interim Analysis (4.1)

Careful conduct of a clinical trial according to the protocol has a major impact on the credibility of the results (see ICH E6). Careful monitoring can ensure that difficulties are noticed early and their occurrence or recurrence minimized.

There are two distinct types of monitoring that generally characterize confirmatory clinical trials sponsored by the pharmaceutical industry. One type of monitoring concerns the oversight of the quality of the trial, while the other type involves breaking the blind to make treatment comparisons (i.e., interim analysis). Both types of trial monitoring, in addition to entailing different staff responsibilities, involve access to different types of trial data and information, and thus different principles apply for the control of potential statistical and operational bias.

For the purpose of overseeing the quality of the trial, the checks involved in trial monitoring may include whether the protocol is being followed, the acceptability of data being accrued, the success of planned accrual targets, the appropriateness of the design assumptions, success in keeping patients in the trials, and so on (see sections IV.B to IV.D). This type of monitoring does not require access to information on comparative treatment effects nor unblinding of data and, therefore, has no impact on Type I error. The monitoring of a trial for this purpose is the responsibility of the sponsor (see ICH E6) and can be carried out by the sponsor or an independent group selected by the sponsor. The period for this type of monitoring usually starts with the selection of the trial sites and ends with the collection and cleaning of the last subject's data.

The other type of trial monitoring (interim analysis) involves the accruing of comparative treatment results. Interim analysis requires unblinded (i.e., key breaking) access to treatment group assignment (actual treatment assignment or identification of group assignment) and comparative treatment group summary information. Therefore, the protocol (or appropriate amendments prior to a first analysis) should contain statistical plans for the interim analysis to prevent certain types of bias. This is discussed in sections IV.E and IV.F.

B. Changes in Inclusion and Exclusion Criteria (4.2)

Inclusion and exclusion criteria should remain constant, as specified in the protocol, throughout the period of subject recruitment. Changes may occasionally be appropriate, for example, in long-term trials, where growing medical knowledge either from outside the trial or from interim analyses may suggest a change of entry criteria. Changes may also result from the discovery by monitoring staff that regular violations of the entry criteria are occurring or that seriously low recruitment rates are due to over-restrictive criteria. Changes should be made without breaking the blind and should always be described by a protocol amendment. This amendment should cover any statistical consequences, such as sample size adjustments arising from different event rates, or modifications to the planned analysis, such as stratifying the analysis according to modified inclusion/exclusion criteria.

C. Accrual Rates (4.3)

In trials with a long time-scale for the accrual of subjects, the rate of accrual should be monitored. If it falls appreciably below the projected level, the reasons should be identified and remedial actions taken to protect the power of the trial and alleviate concerns about selective entry and other aspects of quality. In a multicenter trial, these considerations apply to the individual centers.

D. Sample Size Adjustment (4.4)

In long-term trials there will usually be an opportunity to check the assumptions which underlie the original design and sample size calculations. This may be particularly important if the trial specifications have been made on preliminary and/or uncertain information. An interim check conducted on the blinded data may reveal that overall response variances, event rates or survival experience are not as anticipated. A revised sample size may then be calculated using suitably modified assumptions, and should be justified and documented in a protocol amendment and in the clinical study report. The steps taken to preserve blindness and the consequences, if any, for the Type I error and the width of confidence intervals should be explained. The potential need for re-estimation of the sample size should be envisaged in the protocol whenever possible (see section III.E).

E. Interim Analysis and Early Stopping (4.5)

An interim analysis is any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to formal completion of a trial. Because the number, methods, and consequences of these comparisons affect the interpretation of the trial, all interim analyses should be carefully planned in advance and described in the protocol. Special circumstances may dictate the need for an interim analysis that was not defined at the start of a trial. In these cases, a protocol amendment describing the interim analysis should be completed prior to unblinded access to treatment comparison data. When an interim analysis is planned with the intention of deciding whether or not to terminate a trial, this is usually accomplished by the use of a group sequential design that employs statistical monitoring schemes as guidelines (see section III.D). The goal of such an interim analysis is to stop the trial early if the superiority of the treatment under study is clearly established, if the demonstration of a relevant treatment difference has become unlikely, or if unacceptable adverse effects are apparent. Generally, boundaries for monitoring efficacy require more evidence to terminate a trial early (i.e., they are more conservative) than boundaries for monitoring safety. When the trial design and monitoring objective involve multiple endpoints, then this aspect of multiplicity may also need to be taken into account.

The protocol should describe the schedule of interim analyses or, at least, the considerations that will govern its generation, for example, if flexible alpha spending function approaches are to be employed. Further details may be given in a protocol amendment before the time of the first interim analysis. The stopping guidelines and their properties should be clearly described in the protocol or amendments. The potential effects of early stopping on the analysis of other important variables should also be considered. This material should be written or approved by the data monitoring committee (see section IV.F), when the trial has one. Deviations from the planned procedure always bear the potential of invalidating the trial results. If it becomes necessary to make changes to the trial, any consequent changes to the statistical procedures should be specified in an amendment to the protocol at the earliest opportunity, especially discussing the impact on any analysis and inferences that such changes may cause. The procedures selected should always ensure that the overall probability of Type I error is controlled.

The execution of an interim analysis should be a completely confidential process because unblinded data and results are potentially involved. All staff involved in the conduct of the trial should remain blind to the results of such analyses, because of the possibility that their attitudes to the trial will be modified and cause changes in the characteristics of patients to be recruited or biases in treatment comparisons. This principle may be applied to all investigator staff and to staff employed by the sponsor except for those who are directly involved in the execution of the interim analysis. Investigators should be informed only about the decision to continue or to discontinue the trial, or to implement modifications to trial procedures.

Most clinical trials intended to support the efficacy and safety of an investigational product should proceed to full completion of planned sample size accrual; trials should be stopped early only for ethical reasons or if the power is no longer acceptable. However, it is recognized that drug development plans involve the need for sponsor access to comparative treatment data for a variety of reasons, such as planning other trials. It is also

recognized that only a subset of trials will involve the study of serious life-threatening outcomes or mortality which may need sequential monitoring of accruing comparative treatment effects for ethical reasons. In either of these situations, plans for interim statistical analysis should be in place in the protocol or in protocol amendments prior to the unblinded access to comparative treatment data in order to deal with the potential statistical and operational bias that may be introduced.

For many clinical trials of investigational products, especially those that have major public health significance, the responsibility for monitoring comparisons of efficacy and/or safety outcomes should be assigned to an external independent group, often called an independent data monitoring committee (IDMC), a data and safety monitoring board, or a data monitoring committee, whose responsibilities should be clearly described.

When a sponsor assumes the role of monitoring efficacy or safety comparisons and therefore has access to unblinded comparative information, particular care should be taken to protect the integrity of the trial and to manage and limit appropriately the sharing of information. The sponsor should ensure and document that the internal monitoring committee has complied with written standard operating procedures and that minutes of decisionmaking meetings, including records of interim results, are maintained.

Any interim analysis that is not planned appropriately (with or without the consequences of stopping the trial early) may flaw the results of a trial and possibly weaken confidence in the conclusions drawn. Therefore, such analyses should be avoided. If unplanned interim analysis is conducted, the clinical study report should explain why it was necessary and the degree to which blindness had to be broken, and provide an assessment of the potential magnitude of bias introduced and the impact on the interpretation of the results.

F. Role of Independent Data Monitoring Committee (IDMC) (4.6)

(see sections 1.25 and 5.5.2 of ICH E6)

An IDMC may be established by the sponsor to assess at intervals the progress of a clinical trial, safety data, and critical efficacy variables and recommend to the sponsor whether to continue, modify or terminate a trial. The IDMC should have written operating procedures and maintain records of all its meetings, including interim results; these should be available for review when the trial is complete. The independence of the IDMC is intended to control the sharing of important comparative information and to protect the integrity of the clinical trial from adverse impact resulting from access to trial information. The IDMC is a separate entity from an institutional review board (IRB) or an independent ethics committee (IEC), and its composition should include clinical trial scientists knowledgeable in the appropriate disciplines, including statistics.

When there are sponsor representatives on the IDMC, their role should be clearly defined in the operating procedures of the committee (e.g., covering whether or not they can vote on key issues). Since these sponsor staff would have access to unblinded information, the

procedures should also address the control of dissemination of interim trial results within the sponsor organization.

V. DATA ANALYSIS CONSIDERATIONS

A. Prespecification of the Analysis (5.1)

When designing a clinical trial, the principal features of the eventual statistical analysis of the data should be described in the statistical section of the protocol. This section should include all the principal features of the proposed confirmatory analysis of the primary variable(s) and the way in which anticipated analysis problems will be handled. In the case of exploratory trials, this section could describe more general principles and directions.

The statistical analysis plan (see Glossary) may be written as a separate document to be completed after finalizing the protocol. In this document, a more technical and detailed elaboration of the principal features stated in the protocol may be included (see section VII.A). The plan may include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data. The plan should be reviewed and possibly updated as a result of the blind review of the data (see section VII.A for definition) and should be finalized before breaking the blind. Formal records should be kept of when the statistical analysis plan was finalized as well as when the blind was subsequently broken.

If the blind review suggests changes to the principal features stated in the protocol, these should be documented in a protocol amendment. Otherwise, it should suffice to update the statistical analysis plan with the considerations suggested from the blind review. Only results from analyses envisaged in the protocol (including amendments) can be regarded as confirmatory.

In the statistical section of the clinical study report, the statistical methodology should be clearly described including when in the clinical trial process methodology decisions were made (see ICH E3).

B. Analysis Sets (5.2)

The set of subjects whose data are to be included in the main analyses should be defined in the statistical section of the protocol. In addition, documentation for all subjects for whom trial procedures (e.g., run-in period) were initiated may be useful. The content of this subject documentation depends on detailed features of the particular trial, but at least demographic and baseline data on disease status should be collected whenever possible.

If all subjects randomized into a clinical trial satisfied all entry criteria, followed all trial procedures perfectly with no losses to follow-up, and provided complete data records, then the set of subjects to be included in the analysis would be self-evident. The design

and conduct of a trial should aim to approach this ideal as closely as possible, but, in practice, it is doubtful if it can ever be fully achieved. Hence, the statistical section of the protocol should address anticipated problems prospectively in terms of how these affect the subjects and data to be analyzed. The protocol should also specify procedures aimed at minimizing any anticipated irregularities in study conduct that might impair a satisfactory analysis, including various types of protocol violations, withdrawals and missing values. The protocol should consider ways both to reduce the frequency of such problems and to handle the problems that do occur in the analysis of data. Possible amendments to the way in which the analysis will deal with protocol violations should be identified during the blind review. It is desirable to identify any important protocol violation with respect to the time when it occurred, its cause, and its influence on the trial result. The frequency and type of protocol violations, missing values, and other problems should be documented in the clinical study report and their potential influence on the trial results should be described (see ICH E3).

Decisions concerning the analysis set should be guided by the following principles: (1) To minimize bias and (2) to avoid inflation of Type I error.

1. Full Analysis Set (5.2.1)

The intention-to-treat (see Glossary) principle implies that the primary analysis should include all randomized subjects. Compliance with this principle would necessitate complete follow-up of all randomized subjects for study outcomes. In practice, this ideal may be difficult to achieve, for reasons to be described. In this document, the term *full analysis set* is used to describe the analysis set which is as complete as possible and as close as possible to the intention-to-treat ideal of including all randomized subjects. Preservation of the initial randomization in analysis is important in preventing bias and in providing a secure foundation for statistical tests. In many clinical trials, the use of the full analysis set provides a conservative strategy. Under many circumstances, it may also provide estimates of treatment effects that are more likely to mirror those observed in subsequent practice.

There are a limited number of circumstances that might lead to excluding randomized subjects from the full analysis set, including the failure to satisfy major entry criteria (eligibility violations), the failure to take at least one dose of trial medication, and the lack of any data post randomization. Such exclusions should always be justified. Subjects who fail to satisfy an entry criterion may be excluded from the analysis without the possibility of introducing bias only under the following circumstances:

a. The entry criterion was measured prior to randomization. (i)
b. The detection of the relevant eligibility violations can be made completely objectively. (ii)

c. All subjects receive equal scrutiny for eligibility violations. (This may be difficult to ensure in an open-label study, or even in a double-blind study if the data are unblinded prior to this scrutiny, emphasizing the importance of the blind review.) (iii)

d. All detected violations of the particular entry criterion are excluded. (iv)

In some situations, it may be reasonable to eliminate from the set of all randomized subjects any subject who took no trial medication. The intention-to-treat principle would be preserved despite the exclusion of these patients provided, for example, that the decision of whether or not to begin treatment could not be influenced by knowledge of the assigned treatment. In other situations it may be necessary to eliminate from the set of all randomized subjects any subject without data post randomization. No analysis should be considered complete unless the potential biases arising from these specific exclusions, or any others, are addressed.

When the full analysis set of subjects is used, violations of the protocol that occur after randomization may have an impact on the data and conclusions, particularly if their occurrence is related to treatment assignment. In most respects, it is appropriate to include the data from such subjects in the analysis, consistent with the intention-to-treat principle. Special problems arise in connection with subjects withdrawn from treatment after receiving one or more doses who provide no data after this point, and subjects otherwise lost to follow-up, because failure to include these subjects in the full analysis set may seriously undermine the approach. Measurements of primary variables made at the time of the loss to follow-up of a subject for any reason, or subsequently collected in accordance with the intended schedule of assessments in the protocol, are valuable in this context; subsequent collection is especially important in studies where the primary variable is mortality or serious morbidity. The intention to collect data in this way should be described in the protocol. Imputation techniques, ranging from the carrying forward of the last observation to the use of complex mathematical models, may also be used in an attempt to compensate for missing data. Other methods employed to ensure the availability of measurements of primary variables for every subject in the full analysis set may require some assumptions about the subjects' outcomes or a simpler choice of outcome (e.g., success/failure). The use of any of these strategies should be described and justified in the statistical section of the protocol, and the assumptions underlying any mathematical models employed should be clearly explained. It is also important to demonstrate the robustness of the corresponding results of analysis, especially when the strategy in question could itself lead to biased estimates of treatment effects.

Because of the unpredictability of some problems, it may sometimes be preferable to defer detailed consideration of the manner of dealing with irregularities until the blind review of the data at the end of the trial, and, if so, this should be stated in the protocol.

2. *Per Protocol Set (5.2.2)*

The *per protocol* set of subjects, sometimes described as the *valid cases*, the *efficacy* sample, or the *evaluable subjects* sample, defines a subset of the subjects in the full analysis set who are more compliant with the protocol and is characterized by criteria such as the following:

a. The completion of a certain prespecified minimal exposure to the treatment regimen (i)
b. The availability of measurements of the primary variable(s) (ii)
c. The absence of any major protocol violations, including the violation of entry criteria (iii)

The precise reasons for excluding subjects from the per protocol set should be fully defined and documented before breaking the blind in a manner appropriate to the circumstances of the specific trial.

The use of the per protocol set may maximize the opportunity for a new treatment to show additional efficacy in the analysis, and most closely reflects the scientific model underlying the protocol. However, the corresponding test of the hypothesis and estimate of the treatment effect may or may not be conservative, depending on the trial. The bias, which may be severe, arises from the fact that adherence to the study protocol may be related to treatment and outcome.

The problems that lead to the exclusion of subjects to create the per protocol set, and other protocol violations, should be fully identified and summarized. Relevant protocol violations may include errors in treatment assignment, the use of excluded medication, poor compliance, loss to followup, and missing data. It is good practice to assess the pattern of such problems among the treatment groups with respect to frequency and time to occurrence.

3. *Roles of the Different Analysis Sets (5.2.3)*

In general, it is advantageous to demonstrate a lack of sensitivity of the principal trial results to alternative choices of the set of subjects analyzed. In confirmatory trials, it is usually appropriate to plan to conduct both an analysis of the full analysis set and a per protocol analysis, so that any differences between them can be the subject of explicit discussion and interpretation. In some cases, it may be desirable to plan further exploration of the sensitivity of conclusions to the choice of the set of subjects analyzed. When the full analysis set and the per protocol set lead to essentially the same conclusions, confidence in the trial results is increased, bearing in mind, however, that the need to exclude a substantial proportion of subjects from the per protocol analysis throws some doubt on the overall validity of the trial.

The full analysis set and the per protocol set play different roles in superiority trials (which seek to show the investigational product to be superior) and in equivalence or noninferiority trials (which seek to show the investigational product to be comparable, see section III.C.2). In superiority trials, the full analysis set is used in the primary analysis (apart from exceptional circumstances) because it tends to avoid over-optimistic estimates of efficacy resulting from a per protocol analysis. This is because the noncompliers included in the full analysis set will generally diminish the estimated treatment effect. However, in an equivalence or noninferiority trial, use of the full analysis set is generally not conservative and its role should be considered very carefully.

C. Missing Values and Outliers (5.3)

Missing values represent a potential source of bias in a clinical trial. Hence, every effort should be undertaken to fulfill all the requirements of the protocol concerning the collection and management of data. In reality, however, there will almost always be some missing data. A trial may be regarded as valid, nonetheless, provided the methods of dealing with missing values are sensible, particularly if those methods are predefined in the protocol. Definition of methods may be refined by updating this aspect in the statistical analysis plan during the blind review. Unfortunately, no universally applicable methods of handling missing values can be recommended. An investigation should be made concerning the sensitivity of the results of analysis to the method of handling missing values, especially if the number of missing values is substantial.

A similar approach should be adopted to exploring the influence of outliers, the statistical definition of which is, to some extent, arbitrary. Clear identification of a particular value as an outlier is most convincing when justified medically as well as statistically, and the medical context will then often define the appropriate action. Any outlier procedure set out in the protocol or the statistical analysis plan should be such as not to favor any treatment group a priori. Once again, this aspect of the analysis can be usefully updated during blind review. If no procedure for dealing with outliers was foreseen in the trial protocol, one analysis with the actual values and at least one other analysis eliminating or reducing the outlier effect should be performed and differences between their results discussed.

D. Data Transformation (5.4)

The decision to transform key variables prior to analysis is best made during the design of the trial on the basis of similar data from earlier clinical trials. Transformations (e.g., square root, logarithm) should be specified in the protocol and a rationale provided, especially for the primary variable(s). The general principles guiding the use of transformations to ensure that the assumptions underlying the statistical methods are met are to be found in standard texts; conventions for particular variables have been developed in a number of specific clinical areas. The decision on whether and how to transform a variable should be influenced by the preference for a scale that facilitates clinical interpretation.

Similar considerations apply to other derived variables, such as the use of change from baseline, percentage change from baseline, the *area under the curve* of repeated measures, or the ratio of two different variables. Subsequent clinical interpretation should be carefully considered, and the derivation should be justified in the protocol. Closely related points are made in section II.B.2.

E. Estimation, Confidence Intervals, and Hypothesis Testing (5.5)

The statistical section of the protocol should specify the hypotheses that are to be tested and/or the treatment effects that are to be estimated in order to satisfy the primary objectives of the trial. The statistical methods to be used to accomplish these tasks should be described for the primary (and preferably the secondary) variables, and the underlying statistical model should be made clear. Estimates of treatment effects should be accompanied by confidence intervals, whenever possible, and the way in which these will be calculated should be identified. A description should be given of any intentions to use baseline data to improve precision or to adjust estimates for potential baseline differences, for example, by means of analysis of covariance.

It is important to clarify whether one- or two-sided tests of statistical significance will be used and, in particular, to justify prospectively the use of one-sided tests. If hypothesis tests are not considered appropriate, then the alternative process for arriving at statistical conclusions should be given. The issue of one-sided or two-sided approaches to inference is controversial, and a diversity of views can be found in the statistical literature. The approach of setting Type I errors for one-sided tests at half the conventional Type I error used in two-sided tests is preferable in regulatory settings. This promotes consistency with the two-sided confidence intervals that are generally appropriate for estimating the possible size of the difference between two treatments.

The particular statistical model chosen should reflect the current state of medical and statistical knowledge about the variables to be analyzed as well as the statistical design of the trial. All effects to be fitted in the analysis (for example, in analysis of variance models) should be fully specified, and the manner, if any, in which this set of effects might be modified in response to preliminary results should be explained. The same considerations apply to the set of covariates fitted in an analysis of covariance. (See also section V.G.) In the choice of statistical methods, due attention should be paid to the statistical distribution of both primary and secondary variables. When making this choice (for example between parametric and nonparametric methods), it is important to bear in mind the need to provide statistical estimates of the size of treatment effects together with confidence intervals (in addition to significance tests).

The primary analysis of the primary variable should be clearly distinguished from supporting analyses of the primary or secondary variables. Within the statistical section of the protocol or the statistical analysis plan there should also be an outline of the way in which data other than the primary and secondary variables will be summarized and

reported. This should include a reference to any approaches adopted for the purpose of achieving consistency of analysis across a range of trials, for example, for safety data.

Modeling approaches that incorporate information on known pharmacological parameters, the extent of protocol compliance for individual subjects, or other biologically based data may provide valuable insights into actual or potential efficacy, especially with regard to estimation of treatment effects. The assumptions underlying such models should always be clearly identified, and the limitations of any conclusions should be carefully described.

F. Adjustment of Significance and Confidence Levels (5.6)

When multiplicity is present, the usual frequentist approach to the analysis of clinical trial data may necessitate an adjustment to the Type I error. Multiplicity may arise, for example, from multiple primary variables (see section II.B.2), multiple comparisons of treatments, repeated evaluation over time, and/or interim analyses (see section IV.E). Methods to avoid or reduce multiplicity are sometimes preferable when available, such as the identification of the key primary variable (multiple variables), the choice of a critical treatment contrast (multiple comparisons), and the use of a summary measure such as *area under the curve* (repeated measures). In confirmatory analyses, any aspects of multiplicity that remain after steps of this kind have been taken should be identified in the protocol; adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan.

G. Subgroups, Interactions, and Covariates (5.7)

The primary variable(s) is often systematically related to other influences apart from treatment. For example, there may be relationships to covariates such as age and sex, or there may be differences between specific subgroups of subjects, such as those treated at the different centers of a multicenter trial. In some instances, an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol. Pretrial deliberations should identify those covariates and factors expected to have an important influence on the primary variable(s), and should consider how to account for these in the analysis to improve precision and to compensate for any lack of balance between treatment groups. If one or more factors are used to stratify the design, it is appropriate to account for those factors in the analysis. When the potential value of an adjustment is in doubt, it is often advisable to nominate the unadjusted analysis as the one for primary attention, the adjusted analysis being supportive. Special attention should be paid to center effects and to the role of baseline measurements of the primary variable. It is not advisable to adjust the main analyses for covariates measured after randomization because they may be affected by the treatments.

The treatment effect itself may also vary with subgroup or covariate. For example, the effect may decrease with age or may be larger in a particular diagnostic category of

subjects. In some cases such interactions are anticipated or are of particular prior interest (e.g., geriatrics); hence a subgroup analysis or a statistical model including interactions is part of the planned confirmatory analysis. In most cases, however, subgroup or interaction analyses are exploratory and should be clearly identified as such; they should explore the uniformity of any treatment effects found overall. In general, such analyses should proceed first through the addition of interaction terms to the statistical model in question, complemented by additional exploratory analysis within relevant subgroups of subjects, or within strata defined by the covariates. When exploratory, these analyses should be interpreted cautiously. Any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted.

H. Integrity of Data and Computer Software Validity (5.8)

The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software (both internally and externally written) used both for data management (data entry, storage, verification, correction, and retrieval) and for processing the data statistically. Data management activities should therefore be based on thorough and effective standard operating procedures. The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

VI. EVALUATION OF SAFETY AND TOLERABILITY

A. Scope of Evaluation (6.1)

In all clinical trials, evaluation of safety and tolerability (see Glossary) constitutes an important element. In early phases this evaluation is mostly of an exploratory nature and is only sensitive to frank expressions of toxicity, whereas in later phases the establishment of the safety and tolerability profile of a drug can be characterized more fully in larger samples of subjects. Later phase controlled trials represent an important means of exploring, in an unbiased manner, any new potential adverse effects, even if such trials generally lack power in this respect.

Certain trials may be designed with the purpose of making specific claims about superiority or equivalence with regard to safety and tolerability compared to another drug or to another dose of the investigational drug. Such specific claims should be supported by relevant evidence from confirmatory trials, similar to that necessary for corresponding efficacy claims.

B. Choice of Variables and Data Collection (6.2)

In any clinical trial, the methods and measurements chosen to evaluate the safety and tolerability of a drug will depend on a number of factors, including knowledge of the adverse effects of closely related drugs, information from nonclinical and earlier clinical trials and possible consequences of the pharmacodynamic/ pharmacokinetic properties of

the particular drug, the mode of administration, the type of subjects to be studied, and the duration of the trial. Laboratory tests concerning clinical chemistry and hematology, vital signs, and clinical adverse events (diseases, signs, and symptoms) usually form the main body of the safety and tolerability data. The occurrence of serious adverse events and treatment discontinuations due to adverse events are particularly important to register (see ICH E2A and ICH E3).

Furthermore, it is recommended that a consistent methodology be used for the data collection and evaluation throughout a clinical trial program to facilitate the combining of data from different trials. The use of a common adverse event dictionary is particularly important. This dictionary has a structure that makes it possible to summarize the adverse event data on three different levels: System-organ class, preferred term, or included term (see Glossary). The preferred term is the level on which adverse events usually are summarized, and preferred terms belonging to the same system-organ class could then be brought together in the descriptive presentation of data (see ICH M1).

C. Set of Subjects to Be Evaluated and Presentation of Data (6.3)

For the overall safety and tolerability assessment, the set of subjects to be summarized is usually defined as those subjects who received at least one dose of the investigational drug. Safety and tolerability variables should be collected as comprehensively as possible from these subjects, including type of adverse event, severity, onset, and duration (see ICH E2B). Additional safety and tolerability evaluations may be needed in specific subpopulations, such as females, the elderly (see ICH E7), the severely ill, or those who have a common concomitant treatment. These evaluations may need to address more specific issues (see ICH E3).

All safety and tolerability variables will need attention during evaluation, and the broad approach should be indicated in the protocol. All adverse events should be reported, whether or not they are considered to be related to treatment. All available data in the study population should be accounted for in the evaluation. Definitions of measurement units and reference ranges of laboratory variables should be made with care; if different units or different reference ranges appear in the same trial (e.g., if more than one laboratory is involved), then measurements should be appropriately standardized to allow a unified evaluation. Use of a toxicity grading scale should be prespecified and justified.

The incidence of a certain adverse event is usually expressed in the form of a proportion relating number of subjects experiencing events to number of subjects at risk. However, it is not always self-evident how to assess incidence. For example, depending on the situation, the number of exposed subjects or the extent of exposure (in person-years) could be considered for the denominator. Whether the purpose of the calculation is to estimate a risk or to make a comparison between treatment groups, it is important that the definition is given in the protocol. This is especially important if long-term treatment is planned and a substantial proportion of treatment withdrawals or deaths are expected. For such

situations, survival analysis methods should be considered and cumulative adverse event rates calculated in order to avoid the risk of underestimation.

In situations when there is a substantial background noise of signs and symptoms (e.g., in psychiatric trials), one should consider ways for accounting for this in the estimation of risk for different adverse events. One such method is to make use of the *treatment emergent* (see Glossary) concept in which adverse events are recorded only if they emerge or worsen relative to pretreatment baseline.

Other methods to reduce the effect of the background noise may also be appropriate, such as ignoring adverse events of mild severity or requiring that an event should have been observed at repeated visits to qualify for inclusion in the numerator. Such methods should be explained and justified in the protocol.

D. Statistical Evaluation (6.4)

The investigation of safety and tolerability is a multidimensional problem. Although some specific adverse effects can usually be anticipated and specifically monitored for any drug, the range of possible adverse effects is very large, and new and unforeseeable effects are always possible. Further, an adverse event experienced after a protocol violation, such as use of an excluded medication, may introduce a bias. This background underlies the statistical difficulties associated with the analytical evaluation of safety and tolerability of drugs, and means that conclusive information from confirmatory clinical trials is the exception rather than the rule.

In most trials, the safety and tolerability implications are best addressed by applying descriptive statistical methods to the data, supplemented by calculation of confidence intervals wherever this aids interpretation. It is also valuable to make use of graphical presentations in which patterns of adverse events are displayed both within treatment groups and within subjects.

The calculation of p-values is sometimes useful, either as an aid to evaluating a specific difference of interest or as a *flagging* device applied to a large number of safety and tolerability variables to highlight differences worthy of further attention. This is particularly useful for laboratory data, which otherwise can be difficult to summarize appropriately. It is recommended that laboratory data be subjected to both a quantitative analysis (e.g., evaluation of treatment means) and a qualitative analysis where counting of numbers above or below certain thresholds are calculated.

If hypothesis tests are used, statistical adjustments for multiplicity to quantify the Type I error are appropriate, but the Type II error is usually of more concern. Care should be taken when interpreting putative statistically significant findings when there is no multiplicity adjustment.

In the majority of trials, investigators are seeking to establish that there are no clinically unacceptable differences in safety and tolerability compared with either a comparator drug or a placebo. As is the case for noninferiority or equivalence evaluation of efficacy, the use of confidence intervals is preferred to hypothesis testing in this situation. In this way, the considerable imprecision often arising from low frequencies of occurrence is clearly demonstrated.

E. Integrated Summary (6.5)

The safety and tolerability properties of a drug are commonly summarized across trials continuously during an investigational product's development and, in particular, at the time of a marketing application. The usefulness of this summary, however, is dependent on adequate and well-controlled individual trials with high data quality.

The overall usefulness of a drug is always a question of balance between risk and benefit. In a single trial, such a perspective could also be considered even if the assessment of risk/benefit usually is performed in the summary of the entire clinical trial program (see section VII.B.2).

For more details on the reporting of safety and tolerability, see section 12 of ICH E3.

VII. REPORTING

A. Evaluation and Reporting (7.1)

As stated in the introduction, the structure and content of clinical study reports is the subject of ICH E3. That ICH guidance fully covers the reporting of statistical work, appropriately integrated with clinical and other material. The current section is therefore relatively brief.

During the planning phase of a trial, the principal features of the analysis should have been specified in the protocol as described in section V. When the conduct of the trial is over and the data are assembled and available for preliminary inspection, it is valuable to carry out the blind review of the planned analysis also described in section V. This pre-analysis review, blinded to treatment, should cover, for example, decisions concerning the exclusion of subjects or data from the analysis sets, the checking of possible transformations and definitions of outliers, the addition to the model of important covariates identified in other recent research, and the reconsideration of the use of parametric or nonparametric methods. Decisions made at this time should be described in the report and should be distinguished from those made after the statistician has had access to the treatment codes, as blind decisions will generally introduce less potential for bias. Statisticians or other staff involved in unblinded interim analysis should not participate in the blind review or in making modifications to the statistical analysis plan. When the blinding is compromised by the possibility that treatment-induced effects may be apparent in the data, special care will be needed for the blind review.

Many of the more detailed aspects of presentation and tabulation should be finalized at or about the time of the blind review so that, by the time of the actual analysis, full plans exist for all its aspects including subject selection, data selection and modification, data summary and tabulation, estimation, and hypothesis testing. Once data validation is complete, the analysis should proceed according to the predefined plans; the more these plans are adhered to, the greater the credibility of the results. Particular attention should be paid to any differences between the planned analysis and the actual analysis as described in the protocol, the protocol amendments, or the updated statistical analysis plan based on a blind review of data. A careful explanation should be provided for deviations from the planned analysis.

All subjects who entered the trial should be accounted for in the report, whether or not they are included in the analysis. All reasons for exclusion from analysis should be documented; for any subject included in the full analysis set but not in the per protocol set, the reasons for exclusion from the latter should also be documented. Similarly, for all subjects included in an analysis set, the measurements of all important variables should be accounted for at all relevant time-points.

The effect of all losses of subjects or data, withdrawals from treatment, and major protocol violations on the main analyses of the primary variable(s) should be considered carefully. Subjects lost to followup, withdrawn from treatment, or with a severe protocol violation should be identified and a descriptive analysis of them provided, including the reasons for their loss and its relationship to treatment and outcome.

Descriptive statistics form an indispensable part of reports. Suitable tables and/or graphical presentations should illustrate clearly the important features of the primary and secondary variables and of key prognostic and demographic variables. The results of the main analyses relating to the objectives of the trial should be the subject of particularly careful descriptive presentation. When reporting the results of significance tests, precise p-values (e.g., $p=0.034$) should be reported rather than making exclusive reference to critical values.

Although the primary goal of the analysis of a clinical trial should be to answer the questions posed by its main objectives, new questions based on the observed data may well emerge during the unblinded analysis. Additional and perhaps complex statistical analysis may be the consequence. This additional work should be strictly distinguished in the report from work which was planned in the protocol.

The play of chance may lead to unforeseen imbalances between the treatment groups in terms of baseline measurements not predefined as covariates in the planned analysis but having some prognostic importance nevertheless. This is best dealt with by showing that an additional analysis which accounts for these imbalances reaches essentially the same conclusions as the planned analysis. If this is not the case, the effect of the imbalances on the conclusions should be discussed.

In general, sparing use should be made of unplanned analyses. Such analyses are often carried out when it is thought that the treatment effect may vary according to some other factor or factors. An attempt may then be made to identify subgroups of subjects for whom the effect is particularly beneficial. The potential dangers of over-interpretation of unplanned subgroup analyses are well known (see also section V.G) and should be carefully avoided. Although similar problems of interpretation arise if a treatment appears to have no benefit or an adverse effect in a subgroup of subjects, such possibilities should be properly assessed and should therefore be reported.

Finally, statistical judgement should be brought to bear on the analysis, interpretation and presentation of the results of a clinical trial. To this end, the trial statistician should be a member of the team responsible for the clinical study report and should approve the clinical report.

B. Summarizing the Clinical Database (7.2)

An overall summary and synthesis of the evidence on safety and efficacy from all the reported clinical trials is required for a marketing application (*expert report* in EU, *integrated summary* reports in the United States, *gaiyou* in Japan). This may be accompanied, when appropriate, by a statistical combination of results.

Within the summary a number of areas of specific statistical interest arise: Describing the demography and clinical features of the population treated during the course of the clinical trial program; addressing the key questions of efficacy by considering the results of the relevant (usually controlled) trials and highlighting the degree to which they reinforce or contradict each other; summarizing the safety information available from the combined database of all the trials whose results contribute to the marketing application; and identifying potential safety issues. During the design of a clinical program, careful attention should be paid to the uniform definition and collection of measurements which will facilitate subsequent interpretation of the series of trials, particularly if they are likely to be combined across trials. A common dictionary for recording the details of medication, medical history and adverse events should be selected and used. A common definition of the primary and secondary variables is nearly always worthwhile and is essential for meta-analysis. The manner of measuring key efficacy variables, the timing of assessments relative to randomization/entry, the handling of protocol violators and deviators, and perhaps the definition of prognostic factors should all be kept compatible unless there are valid reasons not to do so.

Any statistical procedures used to combine data across trials should be described in detail. Attention should be paid to the possibility of bias associated with the selection of trials, to the homogeneity of their results, and to the proper modelling of the various sources of variation. The sensitivity of conclusions to the assumptions and selections made should be explored.

1. Efficacy Data (7.2.1)

Individual clinical trials should always be large enough to satisfy their objectives. Additional valuable information may also be gained by summarizing a series of clinical trials that address essentially identical key efficacy questions. The main results of such a set of trials should be presented in an identical form to permit comparison, usually in tables or graphs that focus on estimates plus confidence limits. The use of meta-analytic techniques to combine these estimates is often a useful addition because it allows a more precise overall estimate of the size of the treatment effects to be generated and provides a complete and concise summary of the results of the trials. Under exceptional circumstances, a meta-analytic approach may also be the most appropriate way, or the only way, of providing sufficient overall evidence of efficacy via an overall hypothesis test. When used for this purpose, the meta-analysis should have its own prospectively written protocol.

2. Safety Data (7.2.2)

In summarizing safety data, it is important to examine the safety database thoroughly for any indications of potential toxicity and to follow up any indications by looking for an associated supportive pattern of observations. The combination of the safety data from all human exposure to the drug provides an important source of information because its larger sample size provides the best chance of detecting the rarer adverse events and, perhaps, of estimating their approximate incidence. However, incidence data from this database are difficult to evaluate because of the lack of a comparator group, and data from comparative trials are especially valuable in overcoming this difficulty. The results from trials which use a common comparator (placebo or specific active comparator) should be combined and presented separately for each comparator providing sufficient data.

All indications of potential toxicity arising from exploration of the data should be reported. The evaluation of the reality of these potential adverse effects should take into account the issue of multiplicity arising from the numerous comparisons made. The evaluation should also make appropriate use of survival analysis methods to exploit the potential relationship of the incidence of adverse events to duration of exposure and/or followup. The risks associated with identified adverse effects should be appropriately quantified to allow a proper assessment of the risk/benefit relationship.

GLOSSARY (Annex 1)

Bayesian approaches: Approaches to data analysis that provide a posterior probability distribution for some parameter (e.g., treatment effect), derived from the observed data and a prior probability distribution for the parameter. The posterior distribution is then used as the basis for statistical inference.

Bias (statistical and operational): The systematic tendency of any factors associated with the design, conduct, analysis and evaluation of the results of a clinical trial to make the estimate of a treatment effect deviate from its true value. Bias introduced through deviations in conduct is referred to as *operational bias*. The other sources of bias listed above are referred to as *statistical bias*.

Blind review: The checking and assessment of data during the period of time between trial completion (the last observation on the last subject) and the breaking of the blind, for the purpose of finalizing the planned analysis.

Content validity: The extent to which a variable (e.g., a rating scale) measures what it is supposed to measure.

Double dummy: A technique for retaining the blind when administering supplies in a clinical trial, when the two treatments cannot be made identical. Supplies are prepared for Treatment A (active and indistinguishable placebo) and for Treatment B (active and indistinguishable placebo). Subjects then take two sets of treatment; either A (active) and B (placebo), or A (placebo) and B (active).

Dropout: A subject in a clinical trial who for any reason fails to continue in the trial until the last visit required of him/her by the study protocol.

Equivalence trial: A trial with the primary objective of showing that the response to two or more treatments differs by an amount which is clinically unimportant. This is usually demonstrated by showing that the true treatment difference is likely to lie between a lower and an upper equivalence margin of clinically acceptable differences.

Frequentist methods: Statistical methods, such as significance tests and confidence intervals, which can be interpreted in terms of the frequency of certain outcomes occurring in hypothetical repeated realizations of the same experimental situation.

Full analysis set: The set of subjects that is as close as possible to the ideal implied by the intention-to-treat principle. It is derived from the set of all randomized subjects by minimal and justified elimination of subjects.

Generalizability, generalization: The extent to which the findings of a clinical trial can be reliably extrapolated from the subjects who participated in the trial to a broader patient population and a broader range of clinical settings.

Global assessment variable: A single variable, usually a scale of ordered categorical ratings, that integrates objective variables and the investigator's overall impression about the state or change in state of a subject.

Independent data monitoring committee (IDMC) (data and safety monitoring board, monitoring committee, data monitoring committee): An independent data monitoring committee that may be established by the sponsor to assess at intervals the progress of a clinical trial, the safety data, and the critical efficacy endpoints, and to recommend to the sponsor whether to continue, modify, or stop a trial.

Intention-to-treat principle: The principle that asserts that the effect of a treatment policy can be best assessed by evaluating on the basis of the intention to treat a subject (i.e., the planned treatment regimen) rather than the actual treatment given. It has the consequence that subjects allocated to a treatment group should be followed up, assessed, and analyzed as members of that group irrespective of their compliance with the planned course of treatment.

Interaction (qualitative and quantitative): The situation in which a treatment contrast (e.g., difference between investigational product and control) is dependent on another factor (e.g., center). A quantitative interaction refers to the case where the magnitude of the contrast differs at the different levels of the factor, whereas for a qualitative interaction the direction of the contrast differs for at least one level of the factor.

Interrater reliability: The property of yielding equivalent results when used by different raters on different occasions.

Intrarater reliability: The property of yielding equivalent results when used by the same rater on different occasions.

Interim analysis: Any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to the formal completion of a trial.

Meta-analysis: The formal evaluation of the quantitative evidence from two or more trials bearing on the same question. This most commonly involves the statistical combination of summary statistics from the various trials, but the term is sometimes also used to refer to the combination of the raw data.

Multicenter trial: A clinical trial conducted according to a single protocol but at more than one site and, therefore, carried out by more than one investigator.

Noninferiority trial: A trial with the primary objective of showing that the response to the investigational product is not clinically inferior to a comparative agent (active or placebo control).

Preferred and included terms: In a hierarchical medical dictionary, for example, the World Health Organization's Adverse Reaction Terminology (WHO-Art), the included term is the lowest level of dictionary term to which the investigator description is coded. The preferred term is the level of grouping of included terms typically used in reporting frequency of occurrence. For example, the investigator text "Pain in the left arm" might be coded to the included term "Joint pain," which is reported at the preferred term level as "Arthralgia."

Per protocol set (valid cases, efficacy sample, evaluable subjects sample): The set of data generated by the subset of subjects who complied with the protocol sufficiently to ensure that these data would be likely to exhibit the effects of treatment according to the underlying scientific model. Compliance covers such considerations as exposure to treatment, availability of measurements, and absence of major protocol violations.

Safety and tolerability: The safety of a medical product concerns the medical risk to the subject, usually assessed in a clinical trial by laboratory tests (including clinical chemistry and hematology), vital signs, clinical adverse events (diseases, signs and symptoms), and other special safety tests (e.g., electrocardiograms, ophthalmology). The tolerability of the medical product represents the degree to which overt adverse effects can be tolerated by the subject.

Statistical analysis plan: A statistical analysis plan is a document that contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

Superiority trial: A trial with the primary objective of showing that the response to the investigational product is superior to a comparative agent (active or placebo control).

Surrogate variable: A variable that provides an indirect measurement of effect in situations where direct measurement of clinical effect is not feasible or practical.

Treatment effect: An effect attributed to a treatment in a clinical trial. In most clinical trials, the treatment effect of interest is a comparison (or contrast) of two or more treatments.

Treatment emergent: An event that emerges during treatment, having been absent pretreatment, or worsens relative to the pretreatment state.

Trial statistician: A statistician who has a combination of education/training and experience sufficient to implement the principles in this guidance and who is responsible for the statistical aspects of the trial.

Guidance for Industry

E 10 Choice of Control Group and Related Issues in Clinical Trials

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2001
ICH

Guidance for Industry

E 10 Choice of Control Group and Related Issues in Clinical Trials

Additional copies are available from:

Office of Training and Communication
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm

or

Office of Communication, Training and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm
Fax: 1-888-CBERFAX or 301-827-3844
Phone: the Voice Information System at 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2001
ICH

TABLE OF CONTENTS

Guidance for Industry[1]

E10 Choice of Control Group and Related Issues in Clinical Trials

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statutes and regulations.

This guidance is intended to assist applicants in choosing a control group for clinical trials intended to demonstrate the efficacy of a treatment. The guidance also discusses related trial design and conduct issues and describes what trials using each design can demonstrate. This guidance does not address the regulatory requirements of any region.

I. INTRODUCTION (1.0)[2]

The choice of control group is always a critical decision in designing a clinical trial. That choice affects the inferences that can be drawn from the trial, the ethical acceptability of the trial, the degree to which bias in conducting and analyzing the study can be minimized, the types of subjects that can be recruited and the pace of recruitment, the kind of endpoints that can be studied, the public and scientific credibility of the results, the acceptability of the results by regulatory authorities, and many other features of the study, its conduct, and its interpretation.

---

[1] This guidance was developed within the Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, July 2000. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan, and the United States.

[2] Arabic numbers reflect the organizational breakdown in the document endorsed by the ICH Steering Committee at *Step 4* of the ICH process, July 2000.

A. General Scheme and Purpose of Guidance (1.1)

The purpose of this guideance is to describe the general principles involved in choosing a control group for clinical trials intended to demonstrate the efficacy of a treatment and to discuss related trial design and conduct issues. This guidance does not address the regulatory requirements in any region, but describes what trials using each design can demonstrate. The general principles described in this guidance are relevant to any controlled trial but the choice of control group is of particularly critical importance to clinical trials carried out during drug development to demonstrate efficacy. The choice of the control group should be considered in the context of available standard therapies, the adequacy of the evidence to support the chosen design, and ethical considerations.

This guidance first describes the purpose of the control group and the types of control groups commonly employed to demonstrate efficacy. It then discusses the critical design and interpretation issues associated with the use of an active control trial to demonstrate efficacy by showing non-inferiority or equivalence to the control (Section 1.5). There are circumstances in which a finding of non-inferiority cannot be interpreted as evidence of efficacy. Specifically, for a finding of non-inferiority to be interpreted as showing efficacy, the trial needs to have had the ability to distinguish effective from less effective or ineffective treatments.

The guidance then describes trials using each kind of control group in more detail (see sections 2.0-2.5.7) and considers, for each:

- Its ability to minimize bias
- Ethical and practical issues associated with its use
- Its usefulness and the quality of inference in particular situations
- Modifications of study design or combinations with other controls that can resolve ethical, practical, or inferential concerns
- Its overall advantages and disadvantages

Several other ICH guidances are particularly relevant to this guidance:

- E3: Structure and Content of Clinical Study Reports
- E4: Dose-Response Information to Support Drug Registration
- E5: Ethnic Factors
- E6: Good Clinical Practice: Consolidated Guideline
- E8: General Considerations for Clinical Trials
- E9: Statistical Principles for Clinical Trials

Although trials using any of the control groups described and discussed in this guidance may be useful and acceptable in clinical trials that serve as the basis for marketing approval in at least some circumstances, they are not equally appropriate or useful in every case. The general approach to selecting the type of control is outlined in Section 3.0, Figure 1, and Table 1.

Although this guidance is focused primarily on clinical trials intended to assess the efficacy of a treatment, many of the considerations discussed also apply to the assessment of specific safety hypotheses and to safety or efficacy comparisons of two treatments.

B. Purpose of Control Group (1.2)

Control groups have one major purpose: to allow discrimination of patient outcomes (for example, changes in symptoms, signs, or other morbidity) caused by the test treatment from outcomes caused by other factors, such as the natural progression of the disease, observer or patient expectations, or other treatment. The control group experience tells us what would have happened to patients if they had not received the test treatment or if they had received a different treatment known to be effective.

If the course of a disease were uniform in a given patient population, or predictable from patient characteristics such that outcome could be predicted reliably for any given subject or group of subjects, results of treatment could simply be compared with the known outcome without treatment. For example, one could assume that pain would have persisted for a defined time, blood pressure would not have changed, depression would have lasted for a defined time, tumors would have progressed, or the mortality after an acute infarction would have been the same as previously seen. In unusual cases, the course of illness is in fact predictable in a defined population and it may be possible to use a similar group of patients previously studied as a historical control (see section 1.3.5). In most situations, however, a concurrent control group is needed because it is not possible to predict outcome with adequate accuracy or certainty.

A concurrent control group is one chosen from the same population as the test group and treated in a defined way as part of the same trial that studies the test treatment, and over the same period of time. The test and control groups should be similar with regard to all baseline and on-treatment variables that could influence outcome, except for the study treatment. Failure to achieve this similarity can introduce a bias into the study. Bias here (and as used in ICH E9) means the systematic tendency of any aspects of the design, conduct, analysis, and interpretation of the results of clinical trials to make the estimate of a treatment effect deviate from its true value. Randomization and blinding are the two techniques usually used to minimize the chance of such bias and to ensure that the test treatment and control groups are similar at the start of the study and are treated similarly in the course of the study (see ICH E9). Whether a trial design includes these features is a critical determinant of its quality and persuasiveness.

1. Randomization (1.2.1)

Assurance that subject populations are similar in test and control groups is best attained by randomly dividing a single sample population into groups that receive the test or control treatments. Randomization avoids systematic differences between groups with respect to known or unknown baseline variables that could affect outcome. Inability to eliminate systematic differences between treatment groups is a major problem of studies without a concurrent randomized control (see external control trials, section 1.3.5). Randomization also provides a sound basis for statistical inference.

2. *Blinding (1.2.2)*

The groups should not only be similar at baseline, but should be treated and observed similarly during the trial, except for receiving the test and control drug. Clinical trials are often *double-blind* (or *double-masked*), meaning that both subjects and investigators, as well as sponsor or investigator staff involved in the treatment or clinical evaluation of subjects, are unaware of each subject's assigned treatment. Blinding is intended to minimize the potential biases resulting from differences in management, treatment, or assessment of patients, or interpretation of results that could arise as a result of subject or investigator knowledge of the assigned treatment. For example:

- Subjects on active drug might report more favorable outcomes because they expect a benefit or might be more likely to stay in a study if they knew they were on active drug.
- Observers might be less likely to identify and report treatment responses in a no-treatment group or might be more sensitive to a favorable outcome or adverse event in patients receiving active drug.
- Knowledge of treatment assignment could affect vigor of attempts to obtain on-study or follow-up data.
- Knowledge of treatment assignment could affect decisions about whether a subject should remain on treatment or receive concomitant medications or other ancillary therapy.
- Knowledge of treatment assignment could affect decisions as to whether a given subject's results should be included in an analysis.
- Knowledge of treatment assignment could affect choice of statistical analysis.

Blinding is intended to ensure that subjective assessments and decisions are not affected by knowledge of treatment assignment.

C. Types of Controls (1.3)

Control groups in clinical trials can be classified on the basis of two critical attributes: (1) the type of treatment used and (2) the method of determining who will be in the control group. The type of control treatment may be any of the following four: (1) placebo, (2) no treatment, (3) different dose or regimen of the study treatment, or (4) a different active treatment. The principal methods of determining who will be in the control group are by randomization or by selection of a control population separate from the population treated in the trial (external or historical control). This document categorizes control groups into five types. The first four are concurrently controlled (the control group and test groups are chosen from the same population and treated concurrently), usually with random assignment to treatment; they are distinguished by the type of control treatment (listed above) used. External (historical) control groups, regardless of the comparator treatment, are considered together as the fifth type because of serious concerns about the ability of such trials to ensure comparability of test and control groups and their ability to minimize important biases, making this design usable only in unusual circumstances.

It is increasingly common to carry out studies that have more than one type of control group. Each type of control group is appropriate in some circumstances, but none is usable or adequate in every situation. The five types of control group are:

1. *Placebo Concurrent Control (1.3.1)*

In a placebo-controlled trial, subjects are randomly assigned to a test treatment or to an identical-appearing treatment that does not contain the test drug. The treatments may be titrated to effect or tolerance, or may be given at one or more fixed doses. Such trials are almost always double-blind. The name of the control suggests that its purpose is to control for *placebo* effect (improvement in a subject resulting from thinking that he or she is taking a drug), but that is not its only or major benefit. Rather, the placebo control design, by allowing blinding and randomization and including a group that receives an inert treatment, controls for all potential influences on the actual or apparent course of the disease other than those arising from the pharmacologic action of the test drug. These influences include spontaneous change (natural history of the disease and regression to the mean), subject or investigator expectations, the effect of being in a trial, use of other therapy, and subjective elements of diagnosis or assessment. Placebo-controlled trials seek to show a difference between treatments when they are studying effectiveness, but may also seek to show lack of difference (of specified size) in evaluating a safety measurement. In that case, the question of whether the trial could have shown such a difference if there had been one is critical (see section 1.5).

The use of a placebo control group does not imply that the control group is untreated. In many placebo-controlled trials, the new treatment and placebo are each added to a common standard therapy (so-called *add-on studies*, see section 2.1.5.2.1).

2. *No-treatment Concurrent Control (1.3.2)*

In a no treatment-controlled trial, subjects are randomly assigned to test treatment or to no (i.e., absence of) study treatment. The principal difference between this design and a placebo-controlled trial is that subjects and investigators are not blind to treatment assignment. Because of the advantages of double-blind designs, this design is likely to be needed and suitable only when it is difficult or impossible to double-blind (e.g., treatments with easily recognized toxicity) and only when there is reasonable confidence that study endpoints are objective and that the results of the trial are unlikely to be influenced by the factors listed in section 1.2.2. Note that it is often possible to have a blinded evaluator carry out endpoint assessment, even if the overall trial is not double-blind. This is a valuable approach and should always be considered in trials that cannot be blinded, but it does not solve the other problems associated with knowing the treatment assignment (see section 1.2.2).

3. *Dose-response Concurrent Control (1.3.3)*

In a randomized, fixed-dose, dose-response trial, subjects are randomized to one of several fixed-dose groups. Subjects may either be placed on their fixed dose initially or be raised to that dose gradually, but the intended comparison is between the groups on their final dose. Dose-response

trials are usually double-blind. They may include a placebo (zero dose) and/or active control. In a concentration-controlled trial, treatment groups are titrated to several fixed-concentration windows; this type of trial is conceptually similar to a fixed-dose, dose-response trial. In a regimen-controlled trial subjects are randomized to two or more regimens of the study drug (e.g., once vs. twice daily, 3 days vs. 7 days).

4. *Active (Positive) Concurrent Control (1.3.4)*

In an active control (or positive control) trial, subjects are randomly assigned to the test treatment or to an active control treatment. Such trials are usually double-blind, but this is not always possible; many oncology trials, for example, are considered difficult or impossible to blind (see section 1.3.2) because of different regimens, different routes of administration, and different toxicities. Active control trials can have two distinct objectives with respect to showing efficacy: (1) to show efficacy of the test treatment by showing it is as good as a known effective treatment or (2) to show efficacy by showing superiority of the test treatment to the active control. They may also be used with the primary objective of comparing the efficacy and/or safety of the two treatments (see section 1.4). Whether the purpose of the trial is to show efficacy of the new treatment or to compare two treatments, the question of whether the trial would be capable of distinguishing effective from less effective or ineffective treatments is critical (see section 1.5).

5. *External Control (Including Historical Control) (1.3.5)*

An externally controlled trial compares a group of subjects receiving the test treatment with a group of patients external to the study, rather than to an internal control group consisting of patients from the same population assigned to a different treatment. The external control can be a group of patients treated at an earlier time (historical control) or a group treated during the same time period but in another setting. The external control may be defined (a specific group of patients) or non-defined (a comparator group based on general medical knowledge of outcome). Use of this latter comparator is particularly treacherous (such trials are usually considered uncontrolled) because general impressions are so often inaccurate. So-called baseline-controlled studies, in which subjects' status on therapy is compared with status before therapy (e.g., blood pressure, tumor size), have no internal control and are thus uncontrolled or externally controlled (see section 2.5).

6. *Multiple Control Groups (1.3.6)*

As will be described further below (see section 1.5.1), it is often possible and advantageous to use more than one kind of control in a single study, e.g., use of both an active control and placebo. Similarly, trials can use several doses of test drug and several doses of an active control, with or without placebo. This design may be useful for active drug comparisons where the relative potency of the two drugs is not well established, or where the purpose of the trial is to establish relative potency.

D. Purposes of Clinical Trials and Related Issues (1.4)

Two purposes of clinical trials should be distinguished: (1) assessment of the efficacy and/or safety of a treatment and (2) assessment of the relative (comparative) efficacy, safety, risk/benefit relationship or utility of two treatments.

1. *Evidence of Efficacy (1.4.1)*

A trial using any of the control types may demonstrate efficacy of the test treatment by showing that it is superior to the control (placebo, no treatment, and low dose of test drug, active drug). An active control trial may, in addition, demonstrate efficacy in some cases by showing the new treatment to be similar in efficacy to a known effective treatment. This similarity establishes the efficacy of the test treatment, however, only if it can be assumed that the active control was effective under the conditions of the trial, as two treatments would also look similar if neither were effective in the trial (see section 1.5).

Clinical trials designed to demonstrate efficacy of a new drug by showing that it is similar in efficacy to a standard agent have been called *equivalence* trials. Most of these are actually non-inferiority trials, attempting to show that the new drug is not less effective than the control by more than a defined amount, generally called the margin.

2. *Comparative Efficacy and Safety (1.4.2)*

In some cases, the focus of the trial is on the comparison of one treatment with another treatment, not the efficacy of the test drug per se. Depending on the therapeutic area, these trials may be seen as providing information that is important for relative risk/benefit assessment. The active comparator(s) should be acceptable to the region for which the data are intended. It is not necessary to demonstrate superiority to the active comparator, and, depending on the situation, it may not be necessary to show non-inferiority. For example, a less effective treatment could have safety advantages and thus be considered useful.

Even though the primary focus of such a trial is the comparison of treatments, rather than demonstration of efficacy, the cautions described for conducting and interpreting non-inferiority trials need to be taken into account (see section 1.5). Specifically, the ability of the comparative trial to detect a difference between treatments when one exists needs to be established because a trial incapable of distinguishing between treatments that are in fact different cannot provide useful comparative information.

3. *Fairness of Comparisons (1.4.3)*

For the comparative trial to be informative concerning relative safety and/or efficacy, the trial needs to be fair; i.e., the conditions of the trial should not inappropriately favor one treatment over the other. In practice, an active control equivalence or non-inferiority trial offered as evidence of efficacy also almost always needs to provide a fair effectiveness comparison with the control, because any doubt as to whether the control in the study had its usual effect would undermine assurance that the trial had assay sensitivity (see section 1.5). Among aspects of trial

design that could unfairly favor one treatment are choice of dose or patient population and selection and timing of endpoints.

a. Dose (1.4.3.1)

In comparing the test drug with an active control, it is important to choose an appropriate dose and dose regimen of the control and test drugs. In examining the results of a comparison of two treatments, it is important to consider whether an apparently less effective treatment has been used at too low a dose or whether the apparently less well-tolerated treatment has been used at too high a dose. In some cases, to show superior efficacy or safety convincingly it will be necessary to study several doses of the control and perhaps several doses of the test treatment.

b. Patient Population (1.4.3.2)

Selection of subjects for an active control trial can affect outcome; the population studied should be carefully considered in evaluating what the trial has shown. For example, if many subjects in a trial have previously failed to respond to the control treatment, there would be a bias in favor of the new treatment. The results of such a trial could not be generalized to the entire population of previously untreated patients. A finding of superiority of the new treatment, however, still would be evidence of the efficacy of the new treatment in the population studied. In fact, a trial of a new treatment in apparent nonresponders to another treatment, in which the nonresponders are randomized to either the new or failed treatment (so long as this does not place the patients at risk), can provide a demonstration of the value of the new treatment in such nonresponders, a clinically valuable observation.

Similarly, it is sometimes possible to identify patient subsets more or less likely to have a favorable response or to have an adverse response to a particular drug. For example, blacks usually respond poorly to the blood pressure effects of beta blockers and angiotensin- converting enzyme inhibitors, so that a comparison of a new antihypertensive with these drugs in these patients would tend to show superiority of the new drug. It would not be appropriate to conclude that the new drug is generally superior. Again, however, a planned trial in a subgroup, with recognition of its limitations and of what conclusion can properly be drawn, could be informative.

c. Selection and Timing of Endpoints (1.4.3.3)

When two treatments are used for the same disease or condition, they may differentially affect various outcomes of interest in that disease, particularly if they represent different classes or modalities of treatment. Therefore, when comparing them in a clinical trial, the choice and timing of endpoints may favor one treatment or the other. For example, thrombolytics in patients with acute myocardial infarction can reduce mortality but increase hemorrhagic stroke risk. If a new, more pharmacologically active, thrombolytic were compared with an older thrombolytic, the more active treatment might look better if the endpoint were mortality, but worse if the endpoint were a composite of mortality and disabling stroke. Similarly, in comparing two analgesics in the management of dental pain, assigning a particularly heavy weight to pain at

early time points would favor the drug with more rapid onset of effect, while assigning more weight to later time points would favor a drug with a longer duration of effect.

E. Assay Sensitivity (1.5)

Assay sensitivity is a property of a clinical trial defined as the ability to distinguish an effective treatment from a less effective or ineffective treatment. Assay sensitivity is important in any trial but has different implications for trials intended to show differences between treatments (superiority trials) and trials intended to show non-inferiority. If a trial intended to demonstrate efficacy by showing superiority of a test treatment to control lacks assay sensitivity, it will fail to show that the test treatment is superior and will fail to lead to a conclusion of efficacy. In contrast, if a trial is intended to demonstrate efficacy by showing a test treatment to be non-inferior to an active control, but lacks assay sensitivity, the trial may find an ineffective treatment to be non-inferior and could lead to an erroneous conclusion of efficacy.

When two treatments within a trial are shown to have different efficacy (i.e., when one treatment is superior), that finding itself demonstrates that the trial had assay sensitivity. In contrast, a successful non-inferiority trial (i.e., one that has shown non-inferiority), or an unsuccessful superiority trial, generally does not contain such direct evidence of assay sensitivity.

1. *Assay Sensitivity in Non-inferiority or Equivalence Trials (1.5.1)*

The presence of assay sensitivity in a non-inferiority or equivalence trial may be deduced from two determinations:

Historical evidence of sensitivity to drug effects, i.e., that similarly designed trials in the past regularly distinguished effective treatments from less effective or ineffective treatments and

Appropriate trial conduct, i.e., that the conduct of the trial did not undermine its ability to distinguish effective treatments from less effective or ineffective treatments.

Historical evidence of sensitivity to drug effects can, and should, be evaluated before beginning a non-inferiority trial. Specifically, it should be determined that, in the specific therapeutic area under study, appropriately designed and conducted trials that used a specific active treatment, or other treatments with similar effects, reliably showed an effect. Optimally, this is demonstrated by finding that the active treatment intended for use as the active control was reliably found superior to placebo. If this is the case, there is historical evidence of sensitivity to drug effects for similarly designed active control trials (see section 1.5.1.1).

Appropriateness of trial conduct can only be fully evaluated after the active control non-inferiority trial is completed. Not only should the design of the non-inferiority trial be similar to that of previous trials used to determine historical evidence of sensitivity to drug effects (e.g., entry criteria, allowable concomitant therapy); but, in addition, the actual study population entered, the concomitant therapies actually used, etc., should be assessed to ensure that conduct of the study was, in fact, similar to the previous trials. The trial should also be conducted with

high quality (e.g. good compliance, few losses to follow-up). Together with historical evidence of sensitivity to drug effects, appropriate trial conduct (section 1.5.1.2) provides assurance of assay sensitivity in the new active control trial.

The design and conduct of a non-inferiority trial thus involve four critical steps:

Determining that historical evidence of sensitivity to drug effects exists. Without this determination, demonstration of efficacy from a showing of non-inferiority is not possible and should not be attempted.

Designing a trial. Important details of the trial design, e.g., study population, concomitant therapy, endpoints, run-in periods, should adhere closely to the design of the trials used to determine that historical evidence of sensitivity to drug effects exists

Setting a margin. An acceptable non-inferiority margin should be defined, taking into account the historical data and relevant clinical and statistical considerations.

Conducting the trial. The trial conduct should also adhere closely to that of the historical trials and should be of high quality.

a. Historical Evidence of Sensitivity to Drug Effects and Choosing the Non-inferiority Margin (1.5.1.1)

As noted earlier, most active control equivalence trials are really non-inferiority trials intended to establish the efficacy of a new treatment. Analysis of the results of non-inferiority trials is discussed in ICH guidances E9 and E3. Briefly, in such a trial, test and known effective treatments are compared. Prior to the trial, an equivalence or non-inferiority margin, sometimes called *delta*, is selected. This margin is the degree of inferiority of the test treatments to the control that the trial will attempt to exclude statistically. If the confidence interval for the difference between the test and control treatments excludes a degree of inferiority of the test treatment as large as, or larger than, the margin, the test treatment can be declared non-inferior; if the confidence interval includes a difference as large as the margin, the test treatment cannot be declared non-inferior.

The margin chosen for a non-inferiority trial cannot be greater than the *smallest effect size that the active drug would be reliably expected to have* compared with placebo in the setting of the planned trial. If a difference between active control and the new drug favors the control by as much as or more than this margin, the new drug might have no effect at all. Identification of the smallest effect size that the active drug would be reliably expected to have is only possible when there is historical evidence of sensitivity to drug effects and, indeed, identification of the margin is based upon that evidence. The margin generally is identified based on past experience in placebo-controlled trials of adequate design under conditions similar to those planned for the new trial, but could also be supported by dose response or active control superiority studies. Regardless of the control groups used in those earlier studies, the value of interest in determining the margin is the measure of superiority of the active treatment to its control, not uncontrolled

measures such as change from baseline. Note that exactly how to calculate the margin is not described in this document, and there is little published experience on how to do this.

The determination of the margin in a non-inferiority trial is based on both statistical reasoning and clinical judgment, should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative. If this is done properly, a finding that the confidence interval for the difference between new drug and the active control excludes a suitably chosen margin provides assurance that the test drug has an effect greater than zero. In practice, the non-inferiority margin chosen usually will be smaller than that suggested by the smallest expected effect size of the active control because of interest in ensuring that some clinically acceptable effect size (or fraction of the control drug effect) was maintained. For example, it would not generally be considered sufficient in a mortality non-inferiority study to ensure that the test treatment had an effect greater than zero; retention of some substantial fraction of the mortality effect of the control would usually be sought. This would also be true in a trial whose primary focus is the relative effectiveness of a test drug and active control (see section 1.4.2), where it would be usual to seek assurance that the test and control drug were quite similar, not simply that the new drug had any effect at all.

The fact that the choice of the margin to be excluded is based on historical evidence gives the non-inferiority trial an element in common with a historically controlled (externally controlled) trial. The non-inferiority trial design is appropriate and reliable only when the historical estimate of drug effect size can be well supported by reference to the results of previous studies of the control drug. These studies should lead to the conclusion that the active control can consistently be distinguished from placebo in appropriately sized trials of design similar to the proposed trial and should identify an effect size that represents the smallest effect that the control can reliably be expected to have. If placebo-controlled trials of a design similar to the one proposed more than occasionally show no difference between the proposed active control and placebo, and this cannot be explained by some characteristic of the study, only superiority of the test drug would be interpretable.

Whether there is historical evidence of sensitivity to drug effects in any given case is to some degree a matter of judgment. In some cases sensitivity to drug effects is clear from the consistency of results of prior placebo-controlled trials or is obvious because the outcome of treated and untreated disease is very different. For example, in many infectious diseases cure rates on effective treatment far exceed the spontaneous cure rates over the course of a short-term study. There are many conditions, however, in which drugs considered effective cannot regularly be shown superior to placebo in well-controlled trials; and one therefore cannot reliably determine a minimum effect the drug will have in the setting of a specific trial. Such conditions tend to include those in which there is substantial improvement and variability in placebo groups, and/or in which the effects of therapy are small or variable, such as depression, anxiety, dementia, angina, symptomatic congestive heart failure, seasonal allergies, and symptomatic gastroesophageal reflux disease.

In all these cases, there is no doubt that the standard treatments are effective because there are many well-controlled trials of each of these drugs that have shown an effect. Based on available

experience, however, it would be difficult to describe trial conditions in which the drug would reliably have at least a minimum effect (i.e., conditions in which there is historical evidence of sensitivity to drug effects) and that, therefore, could be used to identify an appropriate margin. In some cases, the experience on which the historical evidence of sensitivity to drug effects is based may be of questionable relevance, e.g., if standards of treatment and diagnosis have changed substantially over time (for an example, see section 2.1.7.1). If someone proposing to use an active-control or non-inferiority design cannot provide sufficient support for historical evidence of the sensitivity to drug effects of the study with the chosen non-inferiority margin, a finding of non-inferiority cannot be considered informative with respect to efficacy.

As noted, a determination regarding historical evidence of sensitivity to drug effects applies only to trials of a specific design. For a planned non-inferiority trial to be similarly sensitive to drug effects, it is essential that the trial have critical design characteristics similar to those of the historical trials. These design characteristics include, for example, the entry criteria (severity of medical condition, concomitant illness, method of diagnosis), dose and regimen of control drug, concomitant treatments used, the endpoint measured and timing of assessments, and the use of a washout period to exclude selected patients. When differences in study design characteristics are unavoidable or desirable (e.g. because of technological or therapeutic advances), the implications of any differences for the determination of the presence of historical evidence of sensitivity to drug effects and for choice of margin should be carefully considered.

b. Appropriate Trial Conduct (1.5.1.2)

Even where there is historical evidence of sensitivity to drug effects and the new study is similar in design to the past studies, assay sensitivity can be undermined by the actual conduct of the trial. To ensure assay sensitivity of a trial, its conduct should be of high quality and the patients actually enrolled, the treatments (other than the test treatment) actually given, and the assessments actually made should be similar to those of the trials on which the determination of historical sensitivity to drug effects was based.

There are many factors in the conduct of a trial that can reduce the observed difference between an effective treatment and a less effective or ineffective treatment and therefore may reduce a trial's assay sensitivity, such as:

1. Poor compliance with therapy
2. Poor responsiveness of the enrolled study population to drug effects
3. Use of concomitant non-protocol medication or other treatment that interferes with the test drug or that reduces the extent of the potential response
4. An enrolled population that tends to improve spontaneously, leaving no room for further drug-induced improvement
5. Poorly applied diagnostic criteria (patients lacking the disease to be studied)
6. Biased assessment of endpoint because of knowledge that all patients are receiving a potentially active drug; e.g., a tendency to read blood pressure responses as normalized, potentially reducing the difference between test drug and control

Clinical researchers and trial sponsors intend to perform high-quality trials, and the availability of the Good Clinical Practices guidance (ICH E6) will continue to enhance trial quality. Nonetheless, it should be appreciated that in trials intended to show a difference between treatments there is a strong imperative to use a good trial design and minimize trial errors because many trial imperfections increase the likelihood of failing to show a difference between treatments when one exists. In placebo-controlled trials many efforts are made to improve compliance and increase the likelihood that the patient population will be responsive to drug effects to ensure that an effective treatment will be distinguished from placebo. Nonetheless, in many clinical settings, despite the strong stimulus and extensive efforts to ensure trial excellence and assay sensitivity, clinical trials are often unable to reliably distinguish effective drugs from placebo.

In contrast, in trials intended to show that there is not a difference of a particular size (non-inferiority) between two treatments, there may be a much weaker stimulus to engage in many of these efforts to ensure study quality that will help ensure that differences will be detected, i.e., that ensure assay sensitivity. The kinds of trial error that diminish observed differences between treatments (e.g., poor compliance, high placebo response, certain concomitant treatment, misclassification of outcomes) are of particular concern with respect to preservation of assay sensitivity. However, when it is believed that the new drug is actually superior to the control, there will be a strong stimulus to conduct a high quality trial so that the non-inferiority margin is more likely to be excluded. It should also be noted that some kinds of trial errors can increase variance, which would decrease the likelihood of showing non-inferiority by widening the confidence interval, so that a difference between treatment and control greater than the margin could not be excluded. There would therefore be a strong stimulus in non-inferiority trials to reduce such sources of variance as poor measurement technique.

As noted, to determine that a non-inferiority trial had appropriate trial conduct, its conduct should be reviewed not only for the presence of factors that might obscure differences between treatments but also for factors that might make the trial different from the trials that provided the basis for determining the non-inferiority margin. In particular, it should be determined whether any observed differences in the populations enrolled, the use of concomitant therapies, compliance with therapy, and the extent of, and reasons for, dropping out could adversely affect assay sensitivity. Even when the design and conduct of a trial appear to have been quite similar to those of the trials providing the basis for determining the non-inferiority margin, outcomes with the active control treatment that are visibly atypical (e.g., cure rate in an antibiotic trial that is unusually high or low) can indicate that important differences existed.

2. Assay Sensitivity in Trials Intended to Demonstrate Superiority (1.5.2)

The question of assay sensitivity, although particularly critical in non-inferiority trials, actually arises in any trial that fails to detect a difference between treatments, including a placebo-controlled trial and a dose-response trial. If a treatment fails to show superiority to placebo, for example, it means either that the treatment was ineffective or that the study as designed and conducted was not capable of distinguishing an effective treatment from placebo.

A useful approach to the assessment of assay sensitivity in active control trials and in placebo-controlled trials is the *three-arm trial*, including both placebo and a known active treatment, a trial design with several advantages. Such a trial measures effect size (test drug versus placebo) and allows comparison of test drug and active control in a setting where assay sensitivity is established by the active control versus placebo comparison. (See Section 2.1.5.1.1).

II. DETAILED CONSIDERATION OF TYPES OF CONTROL (2.0)

A. Placebo Control (2.1)

1. Description (See Section 1.3.1) (2.1.1)

In a placebo-controlled trial, subjects are assigned, almost always by randomization, to either a test treatment or to a placebo. A placebo is a *dummy* treatment that appears as identical as possible to the test treatment with respect to physical characteristics such as color, weight, taste and smell, but that does not contain the test drug. Some trials may study more than one dose of the test treatment or include both an active control and placebo. In these cases, it may be easier for the investigator to use more than one placebo (*double-dummy*) than to try to make all treatments look the same. The use of placebo facilitates, and is almost always accompanied by, double-blinding (or double-masking). The difference in outcome between the active treatment and placebo groups is the measure of treatment effect under the conditions of the trial. Within this general description there are a wide variety of designs that can be used successfully: Parallel or crossover designs (see ICH E9), single fixed dose or titration in the active drug group, several fixed doses. Several designs meriting special attention will be described below. Note that not every study that includes a placebo is a placebo-controlled study. For example, an active control study could use a placebo for each drug (double-dummy) to facilitate blinding; this is still an active control trial, not a placebo-controlled trial. A placebo-controlled trial is one in which treatment with a placebo is compared with treatment with a test drug.

It should also be noted that not all placebos are completely inactive. For example, some vehicle controls used in studies of topical skin preparations may have beneficial activity. This does not impair the ability of the design to measure the specific effect of the test agent. Special problems arise when the chosen vehicle control may have harmful effects. In this case a *no treatment* arm would allow the measurement of the total effect of the test agent plus its vehicle.

2. Ability to Minimize Bias (2.1.2)

The placebo-controlled trial, using randomization and blinding, generally minimizes subject and investigator bias. Such trials, however, are not impervious to blind-breaking through recognition of pharmacologic effects of one treatment; blinded outcome assessment can enhance bias reduction in such cases. This concern may be particularly relevant in crossover studies.

3. Ethical Issues (2.1.3)

When a new treatment is tested for a condition for which no effective treatment is known, there is usually no ethical problem with a study comparing the new treatment to placebo. Use of a placebo control may raise problems of ethics, acceptability, and feasibility, however, when an effective treatment is available for the condition under study in a proposed trial. In cases where an available treatment is known to prevent serious harm, such as death or irreversible morbidity in the study population, it is generally inappropriate to use a placebo control. There are occasional exceptions, however, such as cases in which standard therapy has toxicity so severe that many patients have refused to receive it.

In other situations, when there is no serious harm, it is generally considered ethical to ask patients to participate in a placebo-controlled trial, even if they may experience discomfort as a result, provided the setting is noncoercive and patients are fully informed about available therapies and the consequences of delaying treatment. Such trials, however, even if ethical, may pose important practical problems. For example, deferred treatment of pain or other symptoms may be unacceptable to patients or physicians and they may not want to participate in a trial that requires this. Whether a particular placebo controlled trial of a new agent will be acceptable to subjects and investigators when there is known effective therapy is a matter of investigator, patient, and institutional review board (IRB)/ independent ethics committee (IEC) judgment, and acceptability may differ among ICH regions. Acceptability could depend on the specific design of the trial and the patient population chosen, as will be discussed below (see section 2.1.5).

Whether a particular placebo-controlled trial is ethical may in some cases depend on what is believed to have been clinically demonstrated under the particular circumstances of the trial. For example, a short term placebo-controlled trial of a new antihypertensive agent in patients with mild essential hypertension and no end-organ disease might be considered generally acceptable, while a longer trial, or one that included sicker patients, probably would not be.

It should be emphasized that use of a placebo or no-treatment control does not imply that the patient does not get any treatment at all. For example, in an oncology trial, when no active drug is approved, patients in both the placebo or no-treatment group and the test drug group will receive needed palliative treatment, such as analgesics, and best supportive care. Many placebo-controlled trials are conducted as *add-on trials*, where all patients receive a specified standard therapy or therapy left to the choice of the treating physician or institution (see section 2.1.5.2.1).

4. Usefulness of Placebo-controlled Trials and Validity of Inference in Particular Situations (2.1.4)

When used to show effectiveness of a treatment, the placebo-controlled trial is as free of assumptions and reliance on external (extra-study) information as it is possible to be. Most problems in the design or conduct of a trial increase the likelihood of failure to demonstrate a treatment difference (and thereby establish efficacy), so that the trial contains built-in incentives for trial excellence. Even when the primary purpose of a trial is comparison of two active agents or assessment of dose-response, the addition of a placebo provides an internal standard that enhances the inferences that can be drawn from the other comparisons.

Placebo-controlled trials also provide the maximum ability to distinguish adverse effects caused by a drug from those resulting from underlying disease or intercurrent illness. Note, however, that when used to show similarity of two treatments, for example, to show that a drug does not have a particular adverse effect by showing similar rates of the event in drug-treated and placebo-treated patients, placebo-controlled trials have the same assay sensitivity problem as any equivalence or non-inferiority trial (see section 1.5.1). To interpret the result, one must know that if the study drug had caused an adverse event, the event would have been observed. Ordinarily, such a study should include an active control treatment that does cause the adverse event in question, but in some cases it may be possible to conclude that a study has assay sensitivity to such an effect by documenting *historical sensitivity to adverse drug effects* for a particular study design.

5. Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Issues (2.1.5)

It is often possible to address the ethical or practical limitations of placebo-controlled trials by using modified study designs that still retain the inferential advantages of these trials. In addition, placebo-controlled trials can be made more informative by including additional treatment groups, such as multiple doses of the test agent or a known active control treatment.

a. Additional Control Groups (2.1.5.1)

i. Three-arm Trial; Placebo and Active Control (2.1.5.1.1)

As noted in section 1.5.1, three-arm trials including an active control as well as a placebo-control group can readily assess whether a failure to distinguish test treatment from placebo implies ineffectiveness of the test treatment or is simply the result of a trial that lacked the ability to identify an active drug. The comparison of placebo to standard drug in such a trial provides internal evidence of assay sensitivity. It is possible to make the active groups larger than the placebo group to improve the precision of the active drug comparison, if this is considered important. This may also make the trial more acceptable to patients and investigators, as there is less chance of being randomized to placebo.

ii. Additional Doses (2.1.5.1.2)

Randomization to several fixed doses of the test drug in addition to placebo allows assessment of dose-response and may be particularly useful in a comparative trial to ensure a fair comparison of treatments (see ICH E4: Dose-Response Information to Support Drug Registration).

iii. Factorial Designs (2.1.5.1.3)

Factorial designs may be used to explore several doses of the investigational drug as monotherapy and in combination with several doses of another agent proposed for use in combination with it. A single study of this type can define the properties of a wide array of combinations. Such studies are common in the evaluation of new antihypertensive therapies, but

can be considered in a variety of settings where more than one treatment is used simultaneously. For example, the independent additive effects of aspirin and streptokinase in preventing mortality after a heart attack were shown in such a trial.

b. Other Modifications of Study Design (2.1.5.2)

i. Add on Study, Placebo-Controlled; Replacement Study. (2.1.5.2.1)

An *add-on* study is a placebo-controlled trial of a new agent conducted in people also receiving standard treatment. Such studies are particularly important when available treatment is known to decrease mortality or irreversible morbidity, and when a non-inferiority trial with standard treatment as the active control cannot be carried out or would be difficult to interpret (see section 1.5). It is common to study anticancer, antiepileptic, and heart failure drugs this way. This design is useful only when standard treatment is not fully effective (which, however, is almost always the case), and it has the advantage of providing evidence of improved clinical outcomes (rather than mere non-inferiority). Efficacy is, of course, established by such studies only for the combination treatment, and the dose in a monotherapy situation might be different from the dose found to be effective in combination. In general, this approach is likely to succeed only when the new and standard treatments possess different pharmacologic mechanisms, although there are exceptions. For example, combination treatments for people with AIDS may show a beneficial effect of pharmacologically related drugs because of delays in development of resistance.

A variation of this design that can sometimes give information on monotherapy, and that is particularly applicable in the setting of chronic disease, is the replacement study, in which the new drug or placebo is added by random assignment to conventional treatment given at an effective dose and the conventional treatment is then withdrawn, usually by tapering. The ability to maintain the subjects' baseline status is then observed in the drug and placebo groups using predefined success criteria. This approach has been used to study steroid-sparing substitutions in steroid-dependent patients, avoiding initial steroid withdrawal and the recrudescence of symptoms in a washout period. The approach has also been used to study antiepileptic drug monotherapy.

ii. Early *Escape*; Rescue Treatment (2.1.5.2.2)

It is possible to design a study to plan for *early escape* from ineffective therapy. Early escape refers to prompt removal of subjects whose clinical status worsens or fails to improve to a defined level (blood pressure not controlled by a prespecified time, seizure rate greater than some prescribed value, blood pressure rising to a certain level, angina frequency above a defined level, liver enzymes failing to normalize by a preset time in patients with hepatitis), who have a single event that treatment was intended to prevent (first recurrence of unstable angina, grand mal seizure, paroxysmal supraventricular arrhythmia), or who otherwise require rescue treatment. In such cases, the need to change treatment becomes a study endpoint. The criteria for deciding whether these endpoints have occurred should be well specified, and the timing of measurements should ensure that patients would not remain untreated with an active drug while their disease is poorly controlled. The primary difficulty with this trial design is that it may give information

only on short-term effectiveness. The randomized withdrawal trial (see section 2.1.5.2.4), however, which can also incorporate early-escape features, can give information on long-term effectiveness.

iii. Limited Placebo Period (2.1.5.2.3)

In a situation where long-term placebo treatment would not be acceptable, the use of a placebo group for a short period at the beginning of an active control trial could establish assay sensitivity (at least for short-term effects). The trial would then continue without the placebo group.

iv. Randomized Withdrawal (2.1.5.2.4)

In a randomized withdrawal trial, subjects receiving a test treatment for a specified time are randomly assigned to continued treatment with the test treatment or to placebo (i.e., withdrawal of active therapy). Subjects for such a trial could be derived from an organized open single-arm study, from an existing clinical cohort (but usually with a protocol-specified *wash-in* phase to establish the initial on-therapy baseline), from the active arm of a controlled trial, or from one or both arms of an active control trial. Any difference that emerges between the group receiving continued treatment and the group randomized to placebo would demonstrate the effect of the active treatment. The pre-randomization observation period on treatment can be of any length; this approach can therefore be used to study long-term persistence of effectiveness when long-term placebo treatment would not be acceptable. The post-withdrawal observation period could be of fixed duration or could use early escape or time to event (e.g., relapse of depression) approaches. As with the early-escape design, careful attention should be paid to procedures for monitoring patients and assessing study endpoints to ensure that patients failing on an assigned treatment are identified rapidly.

The randomized withdrawal approach is useful in several situations. First, it may be suitable for drugs that appear to resolve an episode of recurring illness (e.g., antidepressants), in which case the withdrawal study is, in effect, a relapse-prevention study. Second, it may be used for drugs that suppress a symptom or sign (chronic pain, hypertension, and angina), but where a long-term placebo-controlled trial would be difficult; in this case, the study can establish long-term efficacy. Third, the design is particularly useful in determining how long a therapy should be continued (e.g., post-infarction treatments with a beta-blocker).

The general advantage of randomized withdrawal designs, when used with an early-escape endpoint, such as return of symptoms, is that the period of placebo exposure with poor response that a patient would have to undergo is short.

This type of design can address dosing issues. After all patients have received an initial fixed dose, they could be randomly assigned in the withdrawal phase to several different doses (as well as placebo), a particularly useful approach when there is reason to think the initial and maintenance doses might be different, either on pharmacodynamic grounds or because there is substantial accumulation of active drug resulting from a long half life of parent drug or active metabolite. Note that the randomized withdrawal design could be used to assess dose-response

after an initial placebo-controlled titration study (See ICH E4). The titration study is an efficient design for establishing effectiveness, but does not give good dose-response information in many cases. The randomized withdrawal phase, with responders randomly assigned to several fixed doses and placebo, will permit dose-response to be studied rigorously while allowing the efficiency of the titration design to be used in the initial phase of the trial.

In using randomized withdrawal designs, it is important to appreciate the possibility of withdrawal phenomena, suggesting the wisdom of relatively slow tapering. A patient may develop tolerance to a drug such that no benefit is being accrued, but the drug's withdrawal may lead to disease exacerbation, resulting in an erroneous conclusion of persisting efficacy. It is also important to realize that treatment effects observed in these trials may be larger than those seen in an unselected population because randomized withdrawal studies are *enriched* with responders and exclude people who cannot tolerate the drug. This phenomenon results when the trial explicitly includes only subjects who appear to have responded to the drug or includes only people who have completed a previous phase of study (which is often an indicator of a good response and always indicates ability to tolerate the drug). In the case of studies intended to determine how long a therapy should be continued, such entry criteria provide the study population and comparison of interest.

v. Other Design Considerations (2.1.5.2.5)

In any placebo-controlled study, unbalanced randomization (e.g., 2:1, study drug to placebo) may enhance the safety database and may also make the study more attractive to patients and/or investigators.

6. *Advantages of Placebo-controlled Trials (2.1.6)*

a. Ability to Demonstrate Efficacy (2.1.6.1)

Like other superiority trials, a placebo-controlled trial contains internal evidence of assay sensitivity. When a difference is demonstrated it is interpretable without reference to external findings.

b. Measures *Absolute* Efficacy and Safety (2.1.6.2)

The placebo-controlled trial measures the total pharmacologically mediated effect of treatment. In contrast, an active control trial or a dose-comparison trial measures the effect relative to another treatment. The placebo-controlled trial also allows a distinction between adverse events due to the drug and those due to the underlying disease or *background noise*. The absolute effect size information is valuable in a three-group trial (test, placebo, active), even if the primary purpose of the trial is the test versus active control comparison.

c. Efficiency (2.1.6.3)

Placebo-controlled trials are efficient in that they can detect treatment effects with a smaller sample size than any other type of concurrently controlled study.

d. Minimizing the Effect of Subject and Investigator Expectations (2.1.6.4)

Use of a blinded placebo control may decrease the amount of improvement resulting from subject or investigator expectations because both are aware that some subjects will receive no active drug. This may increase the ability of the study to detect true drug effects.

7. Disadvantages of Placebo-controlled Trials (2.1.7)

a. Ethical Concerns (See Sections 2.1.3 and 2.1.4) (2.1.7.1)

When effective therapy that is known to prevent death or irreversible morbidity exists for a particular population, that population cannot usually be ethically studied in placebo-controlled trials; the particular conditions and populations for which this is true may be controversial. Ethical concerns may also direct studies toward less ill subjects or toward examination of short-term endpoints when long-term outcomes are of greater interest. Where a placebo-controlled trial is unethical and an active control trial would not be credible, it may be very difficult to study new drugs at all. For example, it would not be considered ethical to carry out a placebo-controlled trial of a thrombolytic agent in patients with acute myocardial infarction. Yet it would be difficult in the current environment to establish a valid non-inferiority margin based on historical data because of the emergence of acute revascularization procedures that might alter the size of the benefits of the thrombolytics. The designs described in section 2.1.5 may be useful in some of these cases.

b. Patient and Physician Practical Concerns (2.1.7.2)

Physicians and/or patients may be reluctant to accept the possibility that the patient will be assigned to the placebo treatment, even if there is general agreement that withholding or delaying treatment will not result in harm. Subjects who sense they are not improving may withdraw from treatment because they attribute lack of effect to having been treated with placebo, complicating the analysis of the study. With care, however, withdrawal for lack of effectiveness can sometimes be used as a study endpoint. Although this may provide some information on drug effectiveness, such information is less precise than actual information on clinical status in subjects receiving their assigned treatment.

c. Generalizability (2.1.7.3)

It is sometimes argued that any controlled trial, but especially a placebo-controlled trial, represents an artificial environment that gives results different from true *real world* effectiveness. If study populations are unrepresentative in placebo-controlled trials because of ethical or practical concerns, questions about the generalizability of study results can arise. For example, protocol, investigator, or patient choice from placebo-controlled trials may exclude patients with more serious disease. In some cases, only a limited number of patients or centers may be willing

to participate in studies. Whether these concerns actually (as opposed to theoretically) limit generalizability has not been established.

d. No Comparative Information (2.1.7.4)

Placebo-controlled trials lacking an active control give little useful information about comparative effectiveness, information that is of interest and importance in many circumstances. Such information cannot reliably be obtained from cross-study comparisons, as the conditions of the studies may have been quite different.

B. No-treatment Concurrent Control (See Section 1.3.2) (2.2)

The randomized no-treatment control is similar in its general properties and its advantages and disadvantages to the placebo-controlled trial. Unlike the placebo-controlled trial, however, it cannot be fully blinded, and this can affect all aspects of the trial, including subject retention, patient management, and all aspects of observation (see section 1.2.2). This design is appropriate in circumstances where a placebo-controlled trial would be performed, except that blinding is not feasible or practical. When this design is used, it is desirable to have critical decisions, such as eligibility and endpoint determination or changes in management, made by an observer blinded to treatment assignment. Decisions related to data analysis, such as inclusion of patients in analysis sets, should also be made by individuals without access to treatment assignment. See ICH E9 for further discussion.

C. Dose-response Concurrent Control (See Section 1.3.3) (2.3)

1. Description (2.3.1)

A dose-response study is one in which subjects are randomly assigned to two or more dosage groups, with or without a placebo group. Dose-response studies are carried out to establish the relation between dose and efficacy and adverse effects and/or to demonstrate efficacy. The first use is considered in ICH E4; the use to demonstrate efficacy is the subject of this guidance. Evidence of efficacy could be based on significant differences in pair-wise comparisons between dosage groups or between dosage groups and placebo, or on evidence of a significant positive trend with increasing dose, even if no two groups are significantly different. In the latter case, however, further study may be needed to assess the effectiveness of the low doses. As noted in ICH E9, the particular approach for the primary efficacy analysis should be prespecified.

Studies in which the treatment groups vary in regimen raise many of the same considerations as dose-response trials. Since the use of regimen-controlled trials to establish efficacy is uncommon, the current discussion is focussed on dose-response trials.

There are several advantages to inclusion of a placebo (zero-dose) group in a dose-response study. First, it avoids studies that are uninterpretable because all doses produce similar effects so that one cannot assess whether all doses are equally effective or equally ineffective. Second, the placebo group permits an estimate of the total pharmacologically mediated effect of treatment,

although the estimate may not be very precise if the dosing groups are relatively small. Third, as the drug-placebo difference is generally larger than inter-dose differences, use of placebo may permit smaller sample sizes. The size of various dose groups need not be identical; e.g., larger samples could be used to give more precise information about the effect of smaller doses or be used to increase the power of the study to show a clear effect of what is expected to be the optimal dose. Dose-response studies can include one or more doses of an active control treatment. Randomized withdrawal designs can also assign subjects to multiple dosage levels.

2. *Ability to Minimize Bias (2.3.2)*

If the dose-response study is blinded, it shares with other randomized and blinded designs an ability to minimize subject and investigator bias. When a drug has pharmacologic effects that could break the blind for some patients or investigators, it may be easier to preserve blinding in a dose-response study than in a placebo-controlled trial. Masking treatments may necessitate multiple dummies or preparation of several different doses that look alike.

3. *Ethical Issues (2.3.3)*

The ethical and practical concerns related to a dose-response study are similar to those affecting placebo-controlled trials. Where there is therapy known to be effective in preventing death or irreversible morbidity, it is no more ethically acceptable to randomize deliberately to subeffective control therapy than it is to randomize to placebo. Where therapy is directed at less serious conditions or where the toxicity of the therapy is substantial relative to its benefits, dose-response studies that use lower, potentially less effective and less toxic doses or placebo may be acceptable to patients and investigators.

4. *Usefulness of Dose-response Studies and Validity of Inference in Particular Situations (2.3.4)*

In general, a blinded dose-response study is useful for the determination of efficacy and safety in situations where a placebo-controlled trial would be useful and has similar credibility (see section 2.1.4).

5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Problems (2.3.5)*

In general, the sorts of modification made to placebo-controlled studies to mitigate ethical, practical, or inferential problems are also applicable to dose-response studies (see section 2.1.5).

6. *Advantages of Dose-response Trials (2.3.6)*

a. Efficiency (2.3.6.1)

Although a comparison of a large, fully effective dose to placebo may be maximally efficient for showing efficacy, this design may produce unacceptable toxicity and gives no dose-response

information. When the dose-response is monotonic, the dose-response trial is reasonably efficient in showing efficacy and also yields dose-response information. If the optimally effective dose is not known, it may be more prudent to study a range of doses than to choose a single dose that may prove to be suboptimal or to have unacceptable adverse effects.

b. Possible Ethical Advantage (2.3.6.2)

In some cases, notably those in which there is likely to be dose-related efficacy and dose-related important toxicity, the dose-response study may represent a difference-showing trial that can be ethically or practically conducted, even where a placebo-controlled trial could not be, because there is reason for patients and investigators to accept lesser effectiveness in return for greater safety.

7. Disadvantages of Dose-response Study (2.3.7)

A potential problem that should be recognized is that a positive dose-response trend (i.e., a significant correlation between the dose and the efficacy outcome), without significant pair-wise differences, can establish efficacy (see 2.3.1), but may leave uncertainty as to which doses (other than the largest) are actually effective. Of course, a single-dose study poses a similar problem with respect to doses below the one studied, giving no information at all about such doses.

It should also be appreciated that it is not uncommon to show no difference between doses in a dose-response study; if there is no placebo group this is usually an uninformative outcome.

If the therapeutic range is not known at all, the design may be inefficient, as many patients may be assigned to sub-therapeutic or supratherapeutic doses.

Dose-response designs may be less efficient than placebo-controlled titration designs for showing the presence of a drug effect; they do, however, in most cases provide better dose-response information (see ICH E4).

D. Active Control (See Section 1.3.4) (2.4)

1. Description (2.4.1)

An active control (positive control) trial is one in which an investigational drug is compared with a known active drug. Such trials are randomized and usually double blind. The most crucial design question is whether the trial is intended to show a difference between the two treatments or to show non-inferiority or equivalence. A sponsor intending to demonstrate effectiveness by means of a trial showing non-inferiority of the test drug to a standard agent needs to address the issue of the assay sensitivity of the trial, as discussed in section 1.5. In a non-inferiority or equivalence trial, the active control treatment needs to be of established efficacy at the dose used and under the conditions of the study (see ICH E9: Statistical Principles for Clinical Trials). In general, this means it should be a drug acceptable in the region to which the studies will be submitted for the same indication at the dose being studied. A superiority study favoring the test

drug, on the other hand, is readily interpretable as evidence of efficacy, even if the dose of active control is too low or the active control is of uncertain benefit (but not if it could be harmful). Such a result, however--superiority in the trial of the test agent to the control--is interpretable as true superiority of the test treatment to the control treatment only when the active control is used in appropriate patients at an appropriate dose and schedule (see section 1.4.3). Lack of appropriate use of the control treatment would also make the study unusable as a non-inferiority study, if superiority of the test drug is not shown, because assay sensitivity of the study would not be ensured (see section 1.5.2).

2. *Ability to Minimize Bias (2.4.2)*

A randomized and blinded active control trial generally minimizes subject and investigator bias, but a note of caution is warranted. In a non-inferiority trial, investigators and subjects know that all subjects are getting active drug, although they do not know which one. This could lead to a tendency toward categorizing borderline cases as successes in partially subjective evaluations, e.g., in an antidepressant study, which could decrease observed treatment differences and increase the likelihood that a finding of non-inferiority would not represent evidence of effectiveness.

3. *Ethical Issues (2.4.3)*

Active control trials are generally considered to pose fewer ethical and practical problems than placebo-controlled trials because all subjects receive active treatment. It should be appreciated, however, that subjects receiving a new treatment are not receiving standard therapy (just as a placebo control group is not) and may be receiving an ineffective or harmful drug. This is an important matter if the active control therapy is known to improve survival or decrease the occurrence of irreversible morbidity, i.e. conditions in which a placebo or no treatment control would be unacceptable. There should therefore be a sound rationale for the test treatment. If there is no strong reason to expect the new drug to be at least as good as the standard, an add-on study (see section 2.1.5.2.1) may be more appropriate, if the conditions allow such a design.

4. *Usefulness of Active Control Trials; Validity of Inference in Particular Situations (2.4.4)*

When a new treatment shows an advantage over an active control, the study is readily interpreted as showing efficacy, just as any other superiority trial is, assuming that the active control is not actually harmful. When an active control trial is used to show efficacy by demonstrating non-inferiority, there is the special consideration of assay sensitivity, which is considered above in section 1.5. The active control trial can also be used to assess comparative efficacy if assay sensitivity is established.

5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical, or Inferential Issues (2.4.5)*

As discussed earlier (section 2.1.5), active control trials can include a placebo group, multiple-dose groups of the test drug, and/or other dose groups of the active control. Comparative dose-response studies, in which there are several doses of both test and active control, are typical in analgesic trials. The doses in active control trials can be fixed or titrated, and both cross-over and parallel designs can be used. The assay sensitivity of a non-inferiority trial can sometimes be supported by a randomized placebo-controlled withdrawal phase at the end (see section 2.1.5.2.4) or by an initial short period of comparison to placebo (see section 2.1.5.2.3). Active control superiority studies in selected populations (nonresponders to other therapy or to the active control) can be very useful and are generally easy to interpret, although the results may not be generalizable.

6. *Advantages of Active Control Trials (2.4.6)*

a. Ethical and Practical Advantages (2.4.6.1)

The active control design, whether intended to show non-inferiority or equivalence or superiority, reduces ethical concerns that arise from failure to use drugs with documented important health benefits. It also addresses patient and physician concerns about failure to use documented effective therapy. Recruitment and IRB/IEC approval may be facilitated, and it may be possible to study larger samples. There may be fewer withdrawals due to lack of effectiveness.

b. Information Content (2.4.6.2)

Where superiority to an active treatment is shown, active control studies are readily interpretable regarding evidence of efficacy. The larger sample sizes needed are sometimes more achievable and acceptable in active control trials and can provide more safety information. Active control trials also can, if properly designed, provide information about relative efficacy.

7. *Disadvantages of Active Control Trials (2.4.7)*

a. Information Content (2.4.7.1)

See section 1.5 for discussion of the problem of assay sensitivity and the ability of the trial to support an efficacy conclusion in non-inferiority or equivalence trials. Even when assay sensitivity is supported and the study is suitable for detecting efficacy, there is no direct assessment of effect size, and there is also greater difficulty in quantitating safety outcomes.

b. Large Sample Size (2.4.7.2)

Generally, the non-inferiority margin to be excluded is chosen conservatively in order to be reasonably sure that the margin is not greater than the smallest effect size that the active control would reliably be expected to have. In addition, because there will usually be an intent to rule

out loss of more than some reasonable fraction (see section 1.5.1) of the control drug effect, a still smaller non-inferiority margin is often used. Because the choice of the margin will therefore be conservative, sample sizes may be very large. In an active control superiority trial, the expected difference between two drugs is always smaller than the expected difference between test drug and placebo, again leading to large sample sizes.

E. External Control (Including Historical Control, See Section 1.3.5) (2.5)

1. Description (2.5.1)

An externally controlled trial is one in which the control group consists of patients who are not part of the same randomized study as the group receiving the investigational agent; i.e., there is no concurrently randomized control group. The control group is thus not derived from exactly the same population as the treated population. Usually, the control group is a well-documented population of patients observed at an earlier time (historical control), but it could be a group at another institution observed contemporaneously, or even a group at the same institution but outside the study. An external control study could be a superiority study (e.g. comparison with an untreated group) or a non-inferiority study. Sometimes certain patients from a larger external experience are selected as a control group on the basis of particular characteristics that make them similar to the treatment group; there may even be an attempt to match particular control and treated patients.

In so-called baseline-controlled studies, the patient's state over time is compared with their baseline state. Although these studies are sometimes thought to use *the patient as his own control*, they do not in fact have an internal control. Rather, changes from baseline are compared with an estimate of what would have happened to the patients in the absence of treatment with the test drug. Both baseline-controlled trials and trials that use a more complicated sequential on-off-on (drug, placebo, drug) design, but that do not include a concurrently randomized control group, are of this type. As noted, in these trials the observed changes from baseline or between study periods are always compared, at least implicitly, to some estimate of what would have happened without the intervention. Such estimates are generally made on the basis of general knowledge, without reference to a specific control population. Although in some cases this is plainly reasonable, e.g., when the effect is dramatic, occurs rapidly following treatment, and is unlikely to have occurred spontaneously (e.g., general anesthesia, cardioversion, measurable tumor shrinkage), in most cases it is not so obvious and a specific historical experience should be sought. Designers and analysts of such trials need to be aware of the limitations of this type of study and should be prepared to justify its use.

2. Ability to Minimize Bias (2.5.2)

Inability to control bias is the major and well-recognized limitation of externally controlled trials and is sufficient in many cases to make the design unsuitable. It is always difficult, and in many cases impossible, to establish comparability of the treatment and control groups and thus to fulfill the major purpose of a control group (see section 1.2). The groups can be dissimilar with respect to a wide range of factors, other than use of the study treatment, that could affect outcome,

including demographic characteristics, diagnostic criteria, stage or severity of disease, concomitant treatments, and observational conditions (such as methods of assessing outcome, investigator expectations). Such dissimilarities can include important but unrecognized prognostic factors that have not been measured. Blinding and randomization are not available to minimize bias when external controls are used. It is well documented that untreated historical-control groups tend to have worse outcomes than an apparently similarly chosen control group in a randomized study, possibly reflecting a selection bias. Control groups in a randomized study need to meet certain criteria to be entered into the study, criteria that are generally more stringent and identify a less sick population than is typical of external control groups. An external control group is often identified retrospectively, leading to potential bias in its selection. A consequence of the recognized inability to control bias is that the potential persuasiveness of findings from externally controlled trials depends on obtaining much more extreme levels of statistical significance and much larger estimated differences between treatments than would be considered necessary in concurrently controlled trials.

The inability to control bias restricts use of the external control design to situations in which the effect of treatment is dramatic and the usual course of the disease highly predictable. In addition, use of external controls should be limited to cases in which the endpoints are objective and the impact of baseline and treatment variables on the endpoint is well characterized.

As noted, the lack of randomization and blinding, and the resultant problems with lack of assurance of comparability of test group and control group, make the possibility of substantial bias inherent in this design and impossible to quantitate. Nonetheless, some approaches to design and conduct of externally controlled trials could lead them to be more persuasive and potentially less biased. A control group should be chosen for which there is detailed information, including, where pertinent, individual patient data regarding demographics, baseline status, concomitant therapy, and course on study. The control patients should be as similar as possible to the population expected to receive the test drug in the study and should have been treated in a similar setting and in a similar manner, except with respect to the study therapy. Study observations should use timing and methodology similar to those used in the control patients. To reduce selection bias, selection of the control group should be made before performing comparative analyses; this may not always be feasible, as outcomes from these control groups may have been published. Any matching on selection criteria or adjustments made to account for population differences should be specified prior to selection of the control and performance of the study. Where no obvious single optimal external control exists, it may be advisable to study multiple external controls, providing that the analytic plan specifies conservatively how each will be used in drawing inferences (e.g., study group should be substantially superior to the most favorable control to conclude efficacy). In some cases, it may be useful to have an independent set of reviewers reassess endpoints in the control group and in the test group in a blinded manner according to common criteria.

3. *Ethical Issues (2.5.3)*

When a drug is intended to treat a serious illness for which there is no satisfactory treatment, especially if the new drug is seen as promising on the basis of theoretical considerations, animal

data, or early human experience, there may be understandable reluctance to perform a comparative study with a concurrent control group of patients who would not receive the new treatment. At the same time, it is not responsible or ethical to carry out studies that have no realistic chance of credibly showing the efficacy of the treatment. It should be appreciated that many promising therapies have had less dramatic effects than expected or have shown no efficacy at all when tested in controlled trials. Investigators may, in these situations, be faced with very difficult judgments. It may be tempting in exceptional cases to initiate an externally controlled trial, hoping for a convincingly dramatic effect, with a prompt switch to randomized trials if this does not materialize.

Alternatively, and generally preferably, in dealing with serious illnesses for which there is no satisfactory treatment, but where the course of the disease cannot be reliably predicted, even the earliest studies should be randomized. This is usually possible when studies are carried out before there is an impression that the therapy is effective. Studies can be monitored by independent data monitoring committees so that dramatic benefit can be detected early. The concurrently controlled trial can detect extreme effects very rapidly and, in addition, can detect modest, but still valuable, effects that would not be credibly demonstrated by an externally controlled trial.

4. *Usefulness of Externally Controlled Trials; Validity of Inference in Particular Situations (2.5.4)*

An externally controlled trial should generally be considered only when prior belief in the superiority of the test therapy to all available alternatives is so strong that alternative designs appear unacceptable and the disease or condition to be treated has a well-documented, highly predictable course. It is often possible, even in these cases, to use alternative, randomized, concurrently controlled designs (see section 2.1.5).

Externally controlled trials are most likely to be persuasive when the study endpoint is objective, when the outcome on treatment is markedly different from that of the external control and a high level of statistical significance for the treatment-control comparison is attained, when the covariates influencing outcome of the disease are well characterized, and when the control closely resembles the study group in all known relevant baseline, treatment (other than study drug), and observational variables. Even in such cases, however, there are documented examples of erroneous conclusions arising from such trials.

When an external control trial is considered, appropriate attention to design and conduct may help reduce bias (see section 2.5.2).

5. *Modifications of Design and Combinations with Other Controls That Can Resolve Ethical, Practical or Inferential Problems (2.5.5)*

The external control design can incorporate elements of randomization and blinding through use of a randomized, placebo controlled withdrawal phase, often with early escape provisions, as described earlier (see section 2.1.5.2.4). The results of the initial period of treatment, in which

subjects who appear to respond are identified and maintained on therapy, are thus *validated* by a rigorous, largely assumption- and bias-free study.

6. *Advantages of Externally Controlled Trials (2.5.6)*

The main advantage of an externally controlled trial is that all patients can receive a promising drug, making the study more attractive to patients and physicians.
The design has some potential efficiencies because all patients are exposed to test drug, of particular importance in rare diseases. However, despite the use of a single treatment group in an externally controlled trial, the estimate of the external control group outcome always should be made conservatively, possibly leading to a larger sample size than would be needed in a placebo-controlled trial. Great caution (e.g., applying a more stringent significance level) is called for because there are likely to be both identified and unidentified or unmeasurable differences between the treatment and control groups, often favoring treatment.

7. *Disadvantages of Externally Controlled Trials (2.5.7)*

The externally controlled study cannot be blinded and is subject to patient, observer, and analyst bias; these are major disadvantages. It is possible to mitigate these problems to a degree, but even the steps suggested in section 2.5.2 cannot resolve such problems fully, as treatment assignment is not randomized and comparability of control and treatment groups at the start of treatment, and comparability of treatment of patients during the trial, cannot be ensured or well assessed. It is well documented that externally controlled trials tend to overestimate efficacy of test therapies. It should be recognized that tests of statistical significance carried out in such studies are less reliable than in randomized trials.

III. CHOOSING THE CONCURRENT CONTROL GROUP (3.0)

Table 1 describes the usefulness of specific types of control groups, and Figure 1 provides a decision tree for choosing among different types of control groups. Although the table and figure focus on the choice of control to demonstrate efficacy, some designs also allow comparisons of test and control agents. The choice of control can be affected by the availability of therapies and by medical practices in specific regions.

The potential usefulness of the principal types of control (placebo, active, and dose-response) in specific situations and for specific purposes is shown in Table 1. The table should be used with the text describing the details of specific circumstances in which potential usefulness can be realized. In all cases, it is presumed that studies are appropriately designed. External controls are a case so distinct that they are not included in the table.

In most cases, evidence of efficacy is most convincingly demonstrated by showing superiority to a concurrent control treatment. If a superiority trial is not feasible or is inappropriate for ethical or practical reasons, and if a defined treatment effect of the active control is regularly seen (e.g.,

as it is for antibiotics in most situations), a non-inferiority or equivalence trial can be used and can be persuasive.

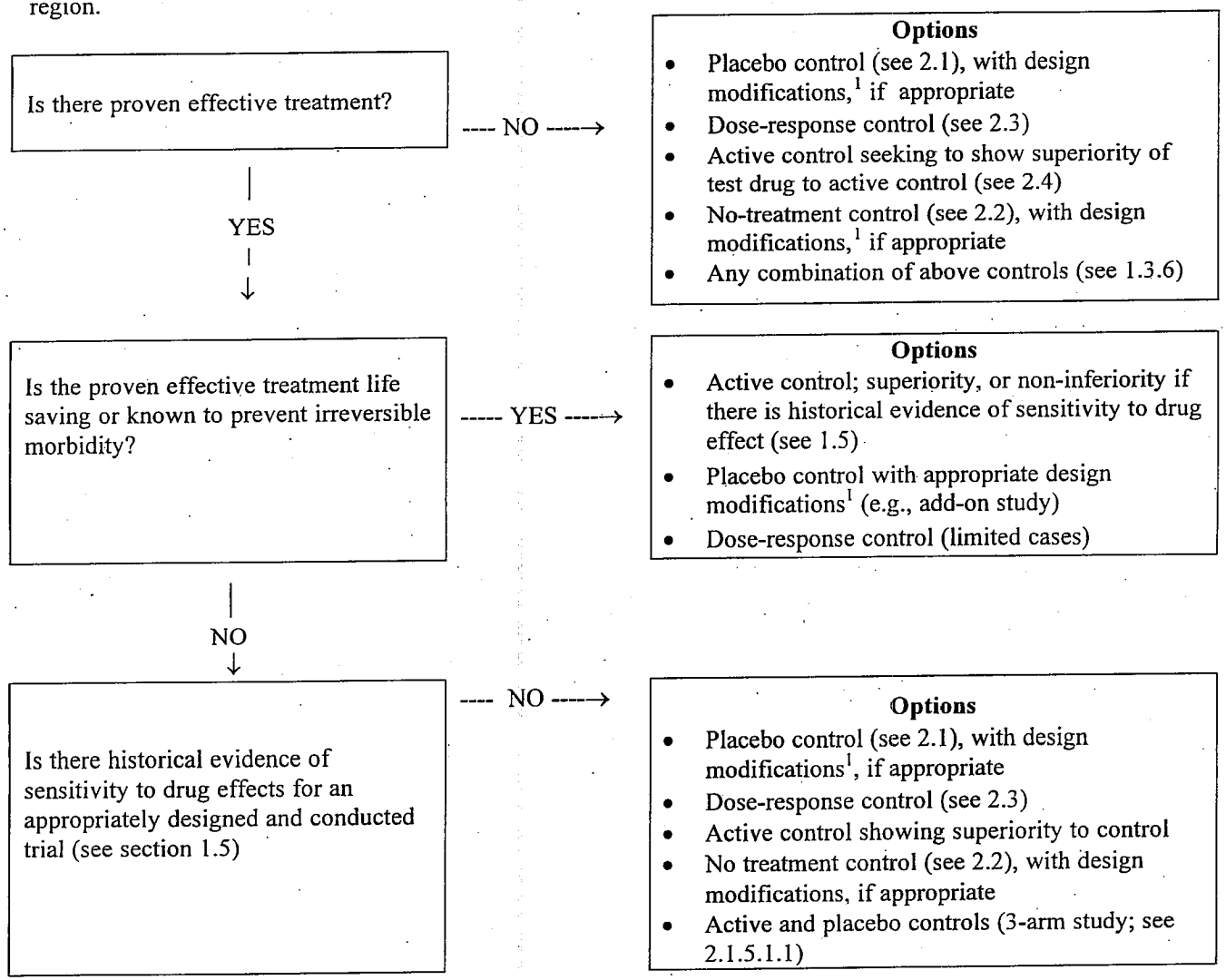
Table 1. Usefulness of Specific Concurrent Control Types in Various Situations

	Type of Control							
Trial Objective	Placebo	Active non-inferiority	Active Superiority	Dose Response (D/R)	Placebo + Active	Placebo + D/R	Active + D/R	Placebo + Active + D/R
Measure *Absolute* effect size	Y	N	N	N	Y	Y	N	Y
Show existence of effect	Y	P	Y	Y	Y	Y	Y	Y
Show Dose-Response relationship	N	N	N	Y	N	Y	Y	Y
Compare therapies	N	P	Y	N	Y	N	P	Y

Y=Yes, N=No, P=Possible, depending on whether there is historical evidence of sensitivity to drug effects

Figure 1: Choosing the Concurrent Control for Demonstrating Efficacy

This figure shows the basic logic for choosing the control group; the decision may depend on the available drugs or medical practices in the specific region.



↓
YES

---------- →

Options

- Placebo control (see 2.1), with design modifications, if appropriate
- Dose-response control
- Active control showing superiority to control
- Active and placebo controls (3-arm study; see 2.1.5.1.1)
- Active control non-inferiority (see 1.5)

[1] Add-on, replacement, early escape, brief placebo period, and randomized withdrawal (see section 2.1.5.2).