ELSEVIER

Antiviral Research 29 (1996) 287–295

Antiviral Research

The influence of signal variation, bias, noise and effect size on statistical significance in treatment studies of the common cold

Jack M. Gwaltney Jr.[a,*], Rose M. Buier[b], James L. Rogers[b]

[a]*Division of Epidemiology and Virology, Department of Internal Medicine, University of Virginia Health Sciences Center, Charlottesville, VA 22908, USA*

[b]*Research and Data Services, Inc., 2037 Bloomingdale Road, Suite 207, Glendale Heights, IL 60139, USA*

Received 21 July 1995; accepted 5 December 1995

Abstract

Many groups are working on new and improved methods of common cold treatment that include antivirals, synthetic viral receptor, compounds which block symptom pathways, and combinations of these approaches. Because the common cold syndrome is in large part subjective, symptom measurement remains an important parameter in evaluating the effectiveness of cold treatments. This review examines the features of the experimental and natural cold testing methods that effect recognition of illness signal and influence its variance and strength. Also, the importance of changes in signal variance and in the magnitude of therapeutic effect size as they relate to statistical probability were compared using a symptom data set from young adults with experimental rhinovirus colds.

Keywords: Acute respiratory infection; Common cold; Effect size; Experimental design; Rhinovirus; Treatment; Variance

1. Introduction

The common cold is a complex upper respiratory tract syndrome caused by multiple different etiologic agents, mostly viruses (Gwaltney, 1990). The clinical features of the illness are in large part subjective and are manifested by respiratory and general symptoms that are widely experienced and recognized. The various groups of respiratory viruses cause somewhat different clinical presentations on average but these are usually not distinguishable at the level of the individual patient (Tyrrell, 1965). Also, variation occurs in the clinical presentation of illness among different individuals with colds due to the same agent. An additional complicating circumstance is the coincidental occurrence of respiratory symptoms due to non-infectious causes.

For these reasons, it can be difficult to recognize treatment effects in randomized, blinded treatment studies of the common cold syndrome. Such studies have historically relied on subjective

* Corresponding author. Tel.: +1 804 924 2093; fax: +1 804 924 0491.

SSDI 0166-3542(95)00935-F

reporting of symptoms, and the problems in their interpretation have been recognized (Smith and Feldman, 1993). This paper examines the causes of signal variance and its relationship to magnitude of effect size in determining statistical significance in cold treatment studies.

2. Materials and methods

2.1. Experimental methods used in cold treatment research

Two methods of evaluating the efficacy and effectiveness of cold treatments have been used—natural cold studies and the virus challenge model. In the natural cold method, volunteers with natural illness are recruited by means of newspaper advertisements and other methods of solicitation. Subjects are usually recruited from free-living populations. In most treatment studies, virologic diagnosis to identify the specific cause of illness has not been attempted. In epidemiologic studies of colds, except for rhinovirus, which accounts for an estimated 40% to 50% of natural colds in adults, no group of respiratory viruses has been associated with more than 5% of illness on an annual basis (Stuart-Harris et al., 1965; Hamre et al., 1966; Gwaltney et al., 1966; Monto and Ullman, 1974). Using currently available methods of viral culture and serology for the different agents, a microbial cause has been found for only 25% to 35% of cases in large longitudinal studies in adults.

In the virus challenge model, subjects are infected with aliquots from the same inoculum pool containing a strain of a single virus (Douglas, 1970; Hendley et al., 1972). Most work has been done with rhinovirus. Viral challenge is performed under standardized conditions of inoculum dose, method of administration, time of day, and season of the year. Subjects are often isolated in individual rooms throughout the duration of the study. Symptom reporting and recording is performed under standardized conditions. The method of symptom recording which has been used most frequently in the past employs the quantification of selected respiratory and systemic symptoms on a 3- or 4-point scale of 'mild', 'moderate', 'severe', ('very severe') and requires a minimum total symptom score for the diagnosis of a cold (Jackson et al., 1958). Treatment is started at a predetermined time after virus challenge. Only data from subjects proven to be infected with the challenge virus are used for analysis of treatment effect. In addition to symptom reporting, the virus challenge model allows measurement of nasal mucus production.

3. Results

3.1. Signal variation

Both the natural cold method and the virus challenge model result in considerable variation in the signal derived from the reporting of symptoms. An example of this with natural rhinovirus infection can be seen in the results of a 3-year longitudinal study of acute respiratory disease in insurance company employees (Gwaltney et al., 1966, 1967). Daily symptom diaries were maintained by the employees who were also tested during colds for infection with rhinovirus and other respiratory viruses. A cold was defined as one or more respiratory symptoms occurring on 2 or more days or two or more respiratory symptoms occurring on 1 or more days. The length of respiratory illnesses of 139 individuals from whom rhinovirus was cultured ranged from 1 to 43 days with a median length of 7 days (Fig. 1A). The mean and standard deviation were 11.3 ± 9.8 days, giving a coefficient of variation of 0.9. The mode, which was 4 days, only represented 12% of the total.

One hundred twenty-one (87%) of the illnesses had durations of 21 days or less. Up to 21 days is the usual duration of rhinovirus shedding during an infection (Cate et al., 1964; Fox and Hall, 1980; Winther et al., 1986). Some of the 18 (13%) illnesses with durations longer than 21 days may have represented cases of secondary bacterial sinusitis, non-infectious conditions that cause respiratory symptoms, or possibly back-to-back infections with different cold viruses. The total number of symptoms per day and the number of

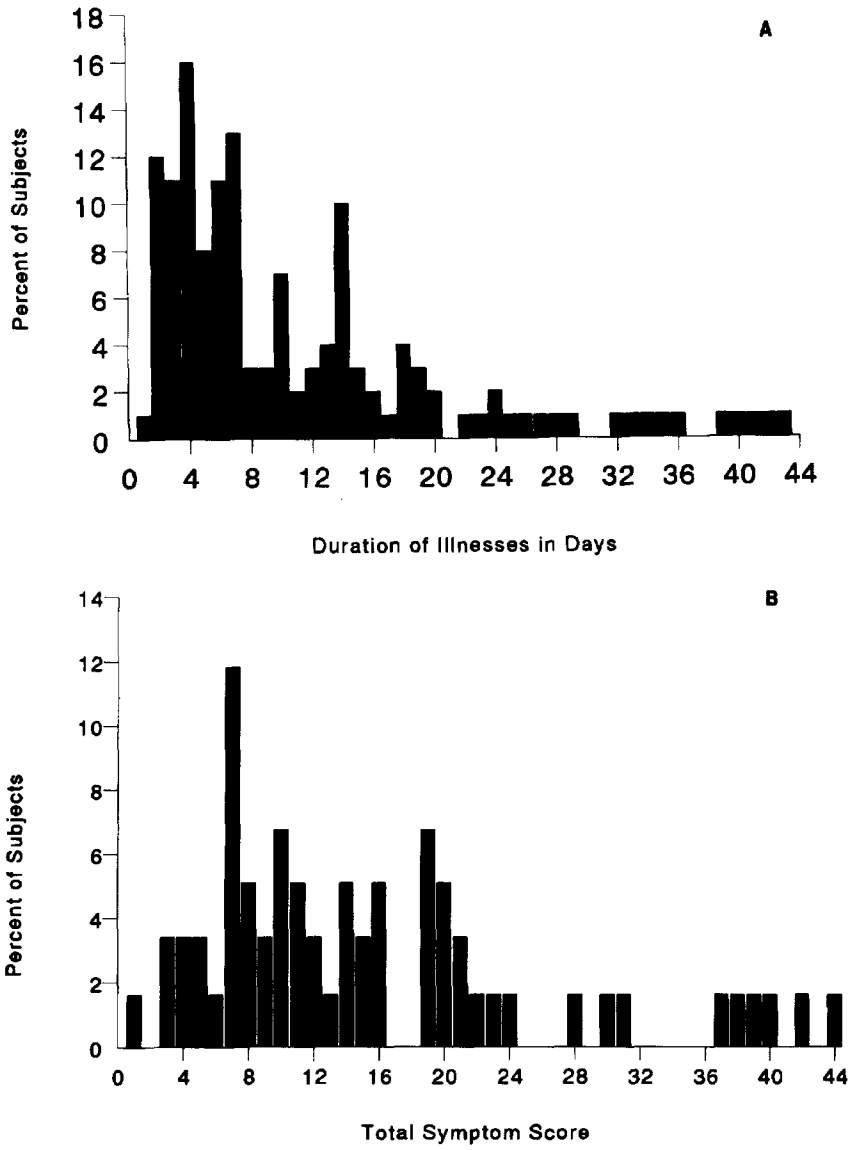

Fig. 1. A. Duration of natural rhinovirus colds in young adults (139 cases). B. Total symptom scores in experimental rhinovirus colds in young adults (58 cases).

different individual symptoms per day also varied considerably. Some persons had asymptomatic infections. In prevalence sampling done during the course of the study, rhinovirus was recovered from 2% of employees who had no respiratory symptoms.

A data set from infected control subjects who were studied by use of the rhinovirus challenge model (Gwaltney et al., 1996) gives another example of the signal variation observed with rhinovirus colds (Fig. 1B). A standard method for quantifying the severity of cold symptoms (Jackson et al., 1958) was used to score the illnesses. The total symptom severity scores of 58 young adults with experimental rhinovirus infection ranged from 1 to 44 during the 5-day observation

period. The mean score was 15.9, and the median score was 13.5. The standard deviation of the mean was ± 10.7 and the coefficient of variation was 0.7. The mode, which was 7, represented only 12% of the group. Individual respiratory symptoms and nasal mucus weights also show a similar, large amount of variation in the challenge model (Parekh et al., 1992).

3.2. Bias and noise

The illness signal sought in cold studies is subject to various known and unknown extraneous influences that can systematically alter its recognition, termed 'bias' (Fletcher et al., 1988). Also, other conditions and events may produce false signals, termed 'noise', that may be mistaken for the signal being sought. Bias and noise can be equally distributed between the test and control groups by the process of randomization. Although this procedure is designed to provide equivalency for bias and noise between treatment and control groups after randomization, their effects on outcome variability persist, interfering with the detection of true signal. Bias and noise, through increased variance, enlarges measurement error and, therefore, interferes with the detection of true signal. The levels of bias and noise can be reduced only by eliminating extraneous influences through alteration of the experiment design to provide more control of essential features.

3.3. Diminishing signal

More importantly, diminishing signal from resolving illness can reduce apparent effect size by exercising a directional impact on the mean symptom score of the control group, forcing the groups together. For example, this may occur if symptom measurements are not started until late in the course of the disease at a time when true signal is diminished due to natural resolution of the illness. It is, therefore, important with a self-limited illness like the common cold to ensure that enrolled subjects have early illnesses.

3.4. The effects of variance and effect size on statistical significance

The previously described symptom data set from 58 subjects with experimental rhinovirus colds (Gwaltney et al., 1996) was used to construct a graphic representation of the interaction of variance and effect size and their influence on statistical significance (Fig. 2). With the amount of variance present in the actual data set, an effect size representing a 40% reduction in mean total symptom score severity in the test group relative to the control group gives a probability of $P = 0.0002$, a highly significant result. Using these data as representative of results obtained with the rhinovirus challenge model, it was possible to examine the impact on statistical significance of holding either the effect size or the variance constant, while systematically altering one or the other of these variables. When the 40% effect size was held constant, the variance could be increased by approximately 250% and still yield a significant probability of $P < 0.05$. In contrast, changing the magnitude of the effect size, while the holding the variance constant, had a much stronger effect on statistical probability. Thus, while controlling the variance to that present in the actual data set, a reduction of the effect size from 40% to 20%, a 50% reduction, no longer yielded a statistically significant result ($P = 0.1$). This suggests that eliminating or reducing extraneous factors that reduce effect size, such as non-compliance in dosing and inaccurate reporting of symptoms due to lack of supervision, is critical in designing cold treatment studies and is more important than restricting variance. It is also apparent that simultaneous increases in variance and decreases in effect size attenuate the significant levels of statistical tests in a non-linear fashion.

3.5. Sources of bias and noise and causes of weak signal in common cold studies

The common cold is usually a short, self-limited illness. In experimental rhinovirus colds, the mean total symptom score and most mean individual symptom scores peak on the second day after

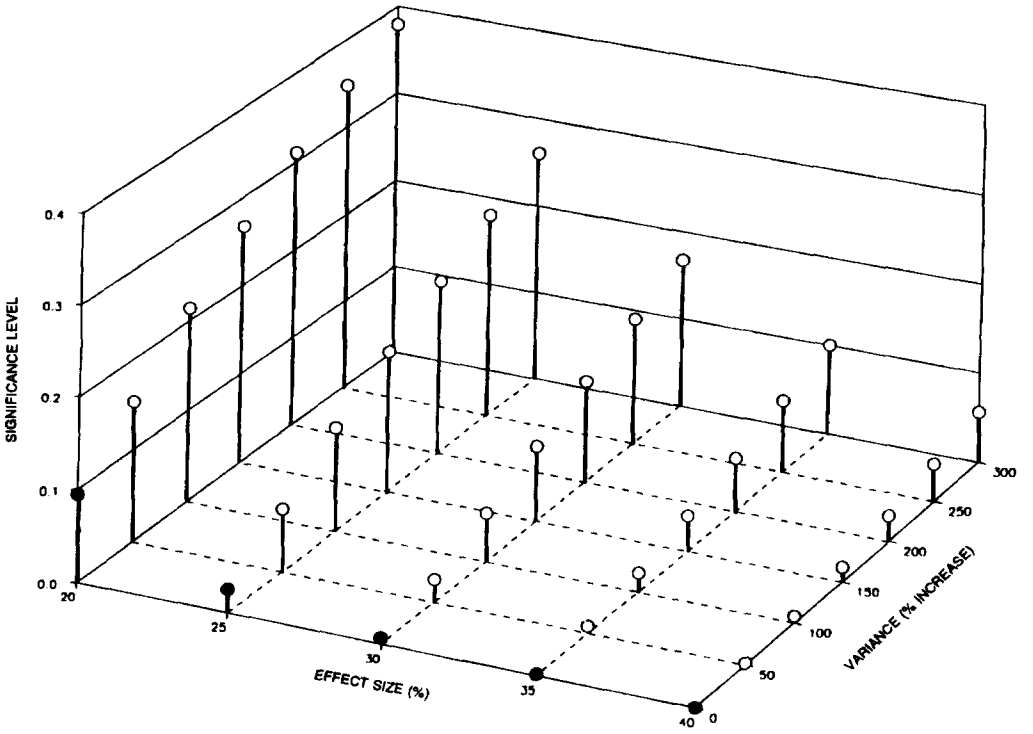

Fig. 2. The influence of signal variance and treatment effect size on significance level in experimental rhinovirus colds. Student's t test was used to evaluate the mean differences in total symptom scores between treatment and control groups. The beginning mean total symptom score (15.9) for the control group and the associated S.D. (var = 10.7) are from part of the data set (58 young adults) of an earlier clinical trial (Gwaltney et al., 1996). Values represented by ● are based on the amount of variance in the actual data set.

viral challenge and then decline over 3–4 days (Parekh et al., 1992). A similar pattern is observed with natural rhinovirus colds (Gwaltney et al., 1967; Rao et al., 1995). Thus, the period leading to maximum symptom development is, on average, approximately 48 h. It is during this brief period that effective rhinovirus cold treatments can be expected to exert their greatest benefit. Thereafter, the natural course of the disease, results in a decline in symptoms in untreated patients, which masks the benefits of effective treatment that are assessed in the recovery phase of the cold. Thus, duration of illness prior to beginning of treatment is a critical variable in common cold treatment studies. It is a potential source of a most important bias and, even if controlled by randomization, may still invalidate a study by weakening the strength of the signal from the control group and reducing the magnitude of the treatment effect that it is possible to measure (Table 1).

Other recognized causes of potential bias and noise in natural cold treatment studies include variability of illness due to different viruses (rhinovirus versus influenza, etc.), the presence of other confounding respiratory illnesses (allergic rhinitis, bacterial pharyngitis, bacterial sinusitis), inaccurate histories given by irresponsible or dishonest subjects, poor compliance in self-medication, and the concomitant use of non-experimental cold treatments. Also, cold treatment studies are subject to the other recognized methodological dangers of clinical trials including various types of enrollment and measurement bias (Fletcher et al., 1988) that, even when randomized, contribute to variance.

Table 1
Features of natural colds compared to experimental colds

Features	Clinical result	Statistical result
Multiple microbial etiology	Larger variation in patterns of illness	Increased variance of signal
	Non-uniformity of response to treatment	Increased variance of signal
Variation in host response to infection with the same agent*	Diversity of individual illness characteristics	Increased variance of signal
Unknown time of infection and variation in the incubation period of different agents	Imprecise clinical recognition of the onset of illness and non-uniformity in the start of treatment, late start of treatment	Increased variance of signal, weak signal
Short duration of illness [a]	Late start of treatment	Weak signal
Other causes of respiratory symptoms [a]	Reporting of false signal	Increased variance due to noise

[a]Also features of experimental infection.

4. Discussion

4.1. Implications for experimental design in cold treatment research

Although variation in symptom severity and duration of illness in individual subjects is one cause of the difficulty encountered in detecting treatment effect in common cold studies, it should be emphasized that this is an inherent characteristic of the signal. In practice, difference in severity of individual illness is not satisfactorily addressed by restricting enrollment to subjects with a predetermined level of severity of illness; i.e., 'mild', 'moderate', 'severe', etc. In both natural and experimental cold studies, the selection of subjects based on the severity of their illness is associated with a reduction, often large, in the size of the samples that can be drawn for study from a target population. More importantly, it is often not possible to predict a cold's overall severity in its early stages; thus, this strategy can delay start of treatment in the crucial, early time period when treatment can be expected to be most effective. As shown in Fig. 2, restricting variance is less important for achieving internal validity than optimizing detectable effect size. Also, such restrictions compromise the external validity of the results of a study. For this reason, it is better to enroll all eligible subjects when they are starting a cold and later, if desired, stratify subjects based on the severity of their illnesses at the time of beginning treatment.

Because signal variation is not an appropriate or attractive target for manipulation in randomized cold treatment studies, experimental design is better directed at ways of eliminating sources of bias and noise. This can be accomplished to some extent by standardization of the infectious process by using the virus challenge model. This model eliminates many of the sources of potential bias and noise present in the natural cold method, and controls for the very important variables of duration of infection and illness prior to starting treatment. This latter problem was encountered in many earlier natural cold treatment studies as shown by the fact that the symptom scores of the subjects in the control groups fell rapidly during the period of early treatment (Howard et al., 1979; Cruthcher and Kantner, 1981; Virtanen, 1983; Gaffey et al., 1988; Berkowitz et al., 1989). For example, in one study (Howard et al., 1979), the mean total cold symptom score in the control group fell from 9.5 at the start of treatment to 6.5 the following day, a 32% reduction. The symptom score in the control group continued to fall at a similar rate on the following day. This indicates that most of the illnesses were already entering the recovery phase at the beginning of the study, and that the period for testing under the most optimal condition to accurately evaluate treatment effect had been missed.

The problem of recruiting subjects with early natural colds may to some extent be avoided, or at least the extent of the problem quantified, by maintaining prospective surveillance with symptom diaries in a population from whom subjects are selected for treatment testing at the first onset of cold symptoms. Experience with this method suggests that early reporting of natural cold onset can be achieved. In one such study, 185 subjects gave a history of cold symptoms of no more than 24 h duration at the time of enrollment into the treatment phase, an inclusion criterion (J.M. Gwaltney Jr., unpublished data). Review of the daily historical record that was obtained prospectively confirmed that only 12 (6%) of the subjects had had illness durations of more than 1 day when they reported for treatment. In another natural cold treatment study with prospective surveillance, the enrollment criterion was an illness of no more than 48 h duration. However, 56 (35%) of 161 subjects had illnesses of 3 or more days duration recorded on their prospective symptom records when they reported for treatment and, thus, did not actually qualify for inclusion (F.G. Hayden et al., unpublished data). From these experiences, it appears that special efforts to encourage early reporting of natural colds can be successful, but even this approach does not always guarantee early reporting of illness. The problem is that although patients are capable of recognizing the onset of new respiratory symptoms, they do not necessarily report them in a timely manner as the onset of an illness.

4.2. Advantages and disadvantages of the virus challenge and natural cold methods

Natural cold studies cannot be expected to determine treatment efficacy in some instances. For example, a drug with antiviral activity against one of the minor cold viruses, such as parainfluenza virus, could not be successfully tested in adults with natural colds since parainfluenza viruses account for, at most, 5% of natural colds in adults. The weak signal from parainfluenza virus colds would be lost in the much larger signal associated with illness due to other viruses.

Also, false signals (noise) are more abundant in natural than in experimental cold studies due to less control of the subject, the infection and the environment. As discussed above, weak signal from the control group may be a problem because of the inability to determine the time of onset of infection and illness, resulting in the study of late colds in which symptom scores are low.

Some of these problems can only be eliminated from the testing situation by using the virus challenge method. However, since the clinical response to experimental infection with rhinovirus shows similar, wide variation to that seen with natural rhinovirus infection, the model offers no advantage in controlling signal variability (Rao et al., 1995). Of more concern is the external validity of the virus challenge model, its ability to predict treatment performance in free-living persons with natural colds. Here, two issues are of particular importance: efficacy relative to the diverse spectrum of microbial agents causing natural common colds and effectiveness for patients living in an unstructured natural setting.

The therapeutic spectrum of a particular cold treatment is dependent on its antiviral and/or pharmacologic properties which are independent of experimental design. The more important epidemiologic question is whether free-living patients can self-diagnose natural colds early enough to start treatment when it will be most effective. Recent rhinovirus challenge studies have shown that the time from intranasal inoculation of virus to the beginning of viral shedding and the onset of symptoms is as brief as 8–12 h (J.M. Harris II et al. 1996). The amount of time required for 1 cycle of viral replication in cell culture is also 8–12 h. By 24 h after experimental viral challenge, colds are well established, although mean severity of illness continues to increase until 48 h after viral inoculation before starting to decline (Rao et al., 1995). Ideally, cold treatments, to have maximal effectiveness, should be started as soon as possible after the onset of illness has been recognized. For rhinovirus colds, this would be within the first 36 h of illness (8–24 h after infection).

In the challenge model, subjects can recognize the onset of experimental rhinovirus cold symp-

toms during this period. As discussed above, trained volunteers keeping prospective daily symptom diaries in a natural free-living environment have illness patterns during natural rhinovirus colds which are similar to those seen with experimental rhinovirus colds (Rao et al., 1995). Also, with training, they are able to recognize colds in the early stages of the illness. Thus, it appears that with the proper education and motivation, patients can be trained to recognize and report early symptoms in natural colds and begin early treatment. If this is correct, then it should be possible to use the natural cold method to confirm the results of the challenge model and validate it as a predictor of effectiveness.

5. Conclusion

This review has focused on the issue of testing treatments for the common cold syndrome, pointing out the problems of signal variation, bias, noise and weak signal, and emphasizing certain advantages of the virus challenge model over the natural cold method. While it is possible that new techniques for viral identification, such as nucleic acid probes, will allow better diagnosis of the microbial etiology of natural colds, the virus challenge model will continue to be an important method for testing cold treatments, especially because of the ability to accurately determine time of onset of infection. In some situations, it may be the only method which is feasible, while in others it may be desirable to correlate virus challenge results with those obtained by the natural cold method. With the natural cold method, it should be recognized that less signal will be produced in late illnesses and that this coupled with greater variance will make signal detection more difficult. Therefore, smaller effect sizes should be expected with the natural cold method than with the virus challenge model. Also, if testing of a rhinovirus cold treatment yields negative results in the virus challenge model, it is unlikely that the treatment will prove effective when tested in patients with natural colds, making the virus challenge model an attractive starting point in the overall study plan.

Acknowledgements

The authors thank Katherine F. Adams, RN and Patsy P. Beasley, RN for data collection; Jennifer Huff for statistical analysis; and Jacqueline Grubbs for manuscript preparation.

References

Berkowitz, R.B., Connell, J.T., Dietz, A.J., Greenstein, S.M. and Tinkelman, D.G. (1989) The effectiveness of the nonsedating antihistamine loratadine plus pseudoephedrine in the symptomatic management of the common cold. Ann. Allergy 63, 336–339.

Cate, T.R., Couch, R.D. and Johnson, K.M. (1964) Studies with rhinoviruses in volunteers: production of illness, effect of naturally acquired antibody, and demonstration of a protective effect not associated with serum antibody. J. Clin. Invest. 43, 56–67.

Cruthcher, J.E. and Kantner, T.R. (1981) The effectiveness of antihistamines in the common cold. J. Clin. Pharmacol. 21, 9–15.

Douglas Jr., R.G. (1970) Pathogenesis of rhinovirus common colds in human volunteers. Ann. Otol., Rhinol., Laryngol. 79, 563.

Fletcher, R.H., Fletcher, S.W. and Wagner, E.H. (1988) Clinical Epidemiology. The Essentials, 2nd edition, Williams and Wilkins, Baltimore.

Fox, J.P. and Hall, C.E. (1980) Viruses in Families. Surveillance of Families as a Key to Epidemiology of Virus Infections. PSG Publishing Co., Littleton.

Gaffey, M.J., Kaiser, D.L. and Hayden, F.G. (1988) Ineffectiveness of oral terfenadine in natural colds: evidence against histamine as a mediator of common cold symptoms. Pediatr. Infect. Dis. J. 7, 223–228.

Gwaltney Jr., J.M. (1990) The common cold. In: G.L. Mandell, R.G Douglas Jr. and J.E. Bennett (Eds), Principles and Practice of Infectious Diseases (3rd edition), pp. 489–493. Churchill Livingstone, New York.

Gwaltney Jr., J.M., Hendley, J.O., Simon, G. and Jordan Jr., W.S. (1966) Rhinovirus infections in an industrial population. I. The occurrence of illness. New Engl. J. Med. 275, 1261–1268.

Gwaltney Jr., J.M., Hendley, J.O., Simon, G. and Jordan Jr., W.S. (1967) Rhinovirus infections in an industrial population. II. Characteristics of illness and antibody response. J. Am. Med. Assoc. 202, 494–500.

Gwaltney Jr., J.M., Park, J., Paul, R.A., Edelman, D.A., O'Connor, R.R. and Turner, R.B. (1996) A randomized controlled trial of clemastine fumarate in experimental rhinovirus cold. Clin. Infect. Dis. 22, 656–662.

Hamre, D., Connelly Jr., A.P. and Procknow, J.J. (1966) Virologic studies of acute respiratory disease in young adults. IV. Virus isolation during four years of surveillance. Am. J. Epidemiol. 83, 238.

Harris II, J.M. and Gwaltney Jr., J.M. (1996) The incubation periods of experimental rhinovirus infection and illness. Clin. Infect. Dis., in press.

Hendley, J.O., Edmondson Jr., W.P. and Gwaltney Jr., J.M. (1972) Relation between naturally acquired immunity and infectivity of two rhinoviruses in volunteers. J. Infect. Dis. 125, 243–248.

Howard Jr., J.C., Kantner, T.R., Lilienfield, L.S., Princiotto, J.V., Krum, R.E., Crutcher, J.E., Belman, M.A. and Danzig, M.R. (1979) Effectiveness of antihistamines in the symptomatic management of the common cold. J. Am. Med. Assoc. 242, 2414–2417.

Jackson, G.G., Dowling, H.F., Spiesman, I.G. and Boand, A.V. (1958) Transmission of the common cold to volunteers under controlled conditions. I. The common cold as a clinical entity. Arch. Intern. Med. 101, 267–278.

Monto, A.S. and Ullman, B.M. (1974) Acute respiratory illness in an American community. The Tecumseh study. J. Am. Med. Assoc. 227, 164–169.

Parekh, H.H., Cragun, K.T., Hayden, F.G., Hendley, J.O. and Gwaltney Jr., J.M., (1992) Nasal mucus weights in experimental rhinovirus infection. Am. J. Rhinol. 63, 107–110.

Rao, S.R., Hendley, J.O., Hayden, F.G. and Gwaltney Jr., J.M. (1995) Symptom expression in natural and experimental rhinovirus colds. Am. J. Rhinol. 9, 49–52.

Smith, M.B.H. and Feldman, W. (1993) Over-the-counter cold medications. A critical review of clinical trials between 1950 and 1991. J. Am. Med. Assoc. 269, 2258–2634.

Stuart-Harris, C.H., Andrewes, C., Andrews, B.E., et al. (1965) A collaborative study of the aetiology of acute respiratory infections in Britain 1961–4. A report of the Medical Research Council Working Party on acute respiratory virus infections. Brit. Med. J. 2, 319–326.

Tyrrell, D.A.J. (1965) Common Colds and Related Diseases. Williams and Wilkins, Baltimore.

Virtanen, A. (1983) Slow release combined preparation (dexchlorpheniramine + pseudoephedrine) for symptomatic treatment of the common cold. J. Laryngol. Otol. 97, 159–163.

Winther, B., Gwaltney Jr., J.M., Mygind, N., Turner, R.D. and Hendley, J.O. (1986) Sites of recovery after point inoculation of the upper airway. J. Am. Med. Assoc. 256, 1763–1767.