and Sure Death, an itinerant history of epidemic disease in Bombay, was published in December 2008 by the Maritime History Society of India. Their last book was the novel *The Quarantine Papers* (Harper Collins India; 404 pp.; January 2010).

Conflict of interest: None declared.

References

[1] Choksy NH. Letter to the Editor: Professor Lustig's plague serum. *Lancet* 1900;**156:**291–92; Reprinted *Int J Epidemiol* 2013;**42**:654–56.

[2] Nathan R. *The Plague in India, 1896, 1897*. Simla: The Government Central Printing Office, 1898.

[3] Condon JK. *The Bombay Plague, being A History of the Progress of Plague in the Bombay Presidency from September 1896 to June 1899*. Bombay: The Education Society's Steam Press, 1900.

[4] Lustig A, Galeotti G. The Prophylactic and Curative Treatment of Plague. *Brit Med J* 1901;**1:**206–08.

[5] Choksy NH. *The Treatment of Plague with Professor Lustig's Serum*. Bombay: Indian Textile and Municipal Journal Company, 1903.

[6] *The Indian Plague Commission: 1898–99:* Bombay: Minutes of Evidence taken before the Indian Plague Commission at the Secretariat.

[7] Haffkine WM. Concerning Inoculation Against Plague and Pneumonia and the Experimental Study of Curative Methods. *J Hyg (Lond)* 1915;**15:**64–101.

[8] Bannerman WB, Haffkine WM. *Serum-therapy of Plague in India*; (Scientific Memoirs by Officers of the Medical and Sanitary Departments of the Government of India; New Series). Calcutta: Government Printing Office, 1905; p. 73.

[9] The Plague. *Br Med J* 1907;**1:**878.

[10] Choksy KBNH. On recent progress in the serum therapy of plague. *Br Med J* 1908;**1:**1282–84.

[11] Choksy NS. Public health in India. *Br Med J* 1923;569.

Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2013;**42**:662–676

doi:10.1093/ije/dyr162

Commentary: The evolution of methods to assess the effects of treatments, illustrated by the development of treatments for diphtheria, 1825–1918

Annick Opinel,[1]* Ulrich Tröhler,[2] Christian Gluud,[3] Gabriel Gachelin,[4] George Davey Smith,[5] Scott Harris Podolsky[6] and Iain Chalmers[7]

[1]Pharmacoepidemiology and Infectious Diseases Unit, Institut Pasteur/UVSQ EA 4499/Inserm U657, 25, rue du Dr Roux, F- 75724 Paris Cedex 15, [2]Institut für Sozial-und Präventivmedizin, Universität Bern, Finkenhubelweg 11, CH-3012 Bern, Switzerland, [3]The Cochrane Hepato-Biliary Group, The Copenhagen Trial Unit, Centre for, Clinical Intervention Research, Department 3344, Rigshospitalet, Copenhagen, University Hospital, Copenhagen, Denmark, [4]Univ. Paris Diderot, Sorbonne Paris Cité, Laboratoire SPHere, UMR 7219 CNRS, F-75205 Paris, France, [5]MRC Centre for Causal Analyses in Observational Epidemiology, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, [6]Center for the History of Medicine, Countway Medical Library, and Department of Global Health and Social Medicine, Harvard Medical School and [7]James Lind Library, Summertown Pavilion, Middle Way, Oxford OX2 7LG, UK

*Corresponding author. Annick Opinel, Unité de pharmaco-épidémiologie et maladies infectieuses, Institut Pasteur, 25 rue du Docteur Roux, F-75015 Paris, France. Email: annick.opinel@pasteur.fr

Accepted 23 May 2011

Numerical methods to assess the effects of medical interventions were introduced during the 18th century[1] and became increasingly sophisticated between the mid-19th and mid-20th centuries. The transition occurred from reports of single cases and case series, controlled, in essence, by unquantified past experience; through quantitative comparisons with historical controls and concurrent controls subject to selection biases; to the adoption of alternation to ensure that like would be compared with like, sometimes using

blinding to reduce observer biases; with increasing recognition of the need to study sufficiently large numbers of patients (see www.jameslindlibrary.org).

To illustrate these developments in research methodology over this period, we have chosen to look at important therapeutic innovations, and how they were evaluated. Treatment of diphtheria is such a case. Until the end of the 19th century, diphtheria was a major killer, mostly of children, but also of adults. However, patients with diphtheria were among the first to benefit from the development and evaluation of specific treatments for infectious diseases,[2] and deaths from the disease fell dramatically in many countries.[3]

The history of diphtheria treatment is characterized by three inventions—tracheotomy, intubation and serum therapy—introduced between the 1840s and the 1890s. The importance of serum therapy was singled out for award of the first Nobel prize in Physiology or Medicine (1901), implying that the beneficial effects of the therapy had been widely accepted. However, uncertainty resulting from the fluctuating severity of diphtheria over time and debate about which interventions actually influenced outcomes meant that the disease provided fertile ground for developing methods for evaluating the effects of new treatments.

Attention has previously focused upon Johannes Fibiger's 1898 report of the use of serum in patients admitted on alternate days to his hospital in Copenhagen,[4] using the experience of patients admitted on other days for comparison. Drawing on studies reported by French, German, American, Danish, Dutch and British researchers, the current article will situate Fibiger's study as a key representative in a series of assessments of therapeutic efficacy during this critical era, thus providing an example of the evaluation methods practised and internationally considered valuable in that era.

Case reports and case series assessing the effects of tracheotomy and tracheal intubation for laryngeal obstruction caused by diphtheria

Diphtheria most often leads to death if left untreated because of obstruction of the upper airways by an adherent membrane. Pierre-Fidèle Bretonneau—who introduced the term diphtheria in 1817—is usually credited with having documented the first successful use of tracheotomy to relieve laryngeal obstruction caused by the disease. After two unsuccessful operations in 1818 and 1820, his third (done in 1825) was successful.[5] Bretonneau's pupil, Emile Trousseau, also had two failed operations (in 1826 and 1828) before succeeding in 1831.[6,7] In 1855, Trousseau reported on the fate of 216 children in whom he had used tracheotomy at the Hôpital des Enfants-Malades in Paris.[8] Of them, 47 children (22%) had survived, a result that he rated as remarkable, given the dire natural history of laryngeal obstruction caused by diphtheria:

> This result is considerable if one thinks about the social conditions of the children brought here, about the deplorable treatment given by the midwives (...), if one thinks about the disastrous conditions of the hospital itself, where children are placed in the middle of the most serious and most different contagions: so that very often, at a time when everything seems to work well after tracheotomy, scarlet fever, measles, cowpox, whooping-cough introduce formidable complications.[8]

Tracheal intubation as an alternative to tracheotomy, an ancient and forgotten practice, was revived in France in 1855 by a surgeon in Lyon, J-F Reybard (1795–1863), who used silver cannulae to perforate the diphtheritic membrane. His method was presented to the Academy of Medicine as a substitute for tracheotomy, to the satisfaction of opponents of the latter.[9] Intubation was later presented to academicians in greater detail by Eugene Bouchut (1818–91) as a method to replace tracheotomy, provided a new type of cannula was used:[10,11]

> One could replace tracheotomy, a difficult and dangerous operation associated with a mortality of 80 to 90% and sometimes more, with a new operation, bloodless, devoid of all danger, [and] as easy to imagine as to accomplish: this is intubation of the trachea.[10]

Bouchut's report was discussed in November 1858 at the Academy of Medicine and Trousseau was asked to examine and report on the method. As tracheal intubation challenged tracheotomy,[7] Trousseau's long report dismissed the former and promoted the latter, drawing attention to the insufficient number of cases (only seven) treated by intubation.[12] Despite continued criticism of tracheotomy, the operation remained dominant, although tracheal intubation remained on the list of hospital practices. When Bouchut was appointed as chief physician at the Hôpital des Enfants-Malades in Paris, he placed intubation on the top of his list of treatments for diphtheria, followed by tonsillectomy (which he described in great detail). Tracheotomy was relegated to be used only after other techniques had failed.[13]

In 1887, an American paediatrician, Joseph O'Dwyer, using an improved technique of tracheal intubation, published a detailed account of 50 patients with croup treated by intubation, 12 (24%) of whom survived.[14] Tracheal intubation was widely accepted in the USA and rapidly superseded tracheotomy as a standard procedure. Although guidelines for performing intubation were widely publicized in France,[7,15,16]

the procedure was not readily accepted in Europe, both because of the influence of prominent physicians, and because of the perceived risks associated with intubation through inflamed tissues. Intubation became a common practice in Europe only after serum therapy had been introduced, with consequent reduction of local inflammation and the risks associated with it.[7]

In attempts to kill the bacteria, measures used in the late 19th century included disinfection of the upper respiratory tract with glycerine and salicylic acid and washing with calomel or with boric and phenolic acid added to water.[17–19] However, the purported effects of these interventions were not quantified, but supported by statements such as 'the membranes were more easily dislodged after such washings'.[16]

Use of historical and concurrent controls to assess the effects of measures to prevent cross-infection with diphtheria and other organisms in hospital

Although tracheotomy and intubation could be lifesaving, patient fatality rate remained high throughout the 19th century. At the end of the century, about half of the children admitted to hospital with diphtheria died,[20] the most seriously affected patients often being infected with streptococci or staphylococci as well.[18] Joseph Grancher (1843–1907), one of the physicians in charge of the Infectious Diseases Service at the Hôpital des Enfants-Malades, had established that diphtheria was not transmitted by circulating air but rather through person-to-person contacts, or contacts with the personal belongings of diphtheria patients. To reduce such super-infection of diphtheria patients with other micro-organisms, as well as to reduce the spread of diphtheritic infection to uninfected children in the hospital, Grancher established a set of guidelines based on the principles of asepsis and isolation techniques that had been adopted in departments of surgery and obstetrics.[22] Rather than proposing a specialist diphtheria hospital, therefore, Grancher's report to the executive ministry responsible for public health (report read and approved at the *Comité consultatif d'hygiène de France* on November 10, 1890) recommended the implementation of rigorous hygiene and asepsis in existing hospitals, as well as measures to limit cross-infection during the transport of patients by the recently established ambulance service.[23]

Accordingly, Grancher reorganized the wards for which he was responsible by surrounding each bed with a 1.2-m high wire gauze screen to minimize movement between beds and by providing each semi-isolated 'cubicle' thus created with individual equipment, sterilized every other day, for food and care. The staff were required to obey very strict asepsis rules when moving from one cubicle to another, washing their hands with mercury sublimate and changing their overalls. Bed clothes were sterilized after each patient had been discharged from hospital.

Grancher claimed that improvements were evident as soon as these new measures had been introduced.[23] In his wards, there had been 19–35 patients with diphtheritic cross-infections out of an average of 500–600 patients per year in the years before the cubicles had been introduced (~3–6%) compared to only one patient (with a dubious diagnosis) out of 575 patients the year following the introduction of the new procedures. Furthermore, in other wards (for measles, surgery and internal medicine) in the Hôpital des Enfants-Malades, which had not been equipped with cubicles, there had been a total of 153 patients of diphtheritic cross-infections (~3%) of about 4000–5000 patients (an estimate based on the average number of patients in those wards in 1887 and 1888). Grancher gives as an example of the frequency of cross-infections, the fact that there had been three diphtheritic cross-infections (6%) during the first 6 months of 1889 out of a total of 47 patients admitted to Husson ward (for patients with chronic diseases), which had not been equipped with cubicles. In contrast with the reduction in diphtheritic cross-infections, no decrease in the spread of measles, which was anyway assumed to be a more transmissable infection, was observed within Grancher's wards.

The first case series describing serum therapy for diphtheria

Diphtheria's effects are caused by a toxin produced by the bacterium *Corynebacterium diphtheriae*. This toxin produces not only diphtheria's effects in the upper respiratory tract, but also later complications, including myocarditis and peripheral neuropathy. These complications and superinfection with other bacterial pathogens (streptococci, in particular), contribute to the serious morbidity and mortality associated with the disease.

In the early 1890s, in Berlin, Emil von Behring and Shibasabura Kitasato developed a serum from a hyper-immune horse, which seemed to confer passive immunity on patients with diphtheria. Experience with this serum was first reported in a paper published in 1893.[24] The authors were cautious in presenting their findings, noting that 'The innocuousness of a treatment . . . is one of the preconditions for justifying a recommendation that it be introduced for therapy in humans. The second, even more important precondition, is evidence of the value foreseen by using it'.

Box 1 Extract from the report by Behring, Boer and Kossel of their use of anti-diphtheritic serum (Source: Behring *et al.*[23])

'We make our medicaments ourselves, we test them ourselves, firstly, not in man, but in animals; we determine ourselves the conditions in which they are innocuous, and the limits within which this is the case; we try out their influence on the course of the human diseases produced in animals, using Robert Koch's methods and when we have determined a specifically curative effect, we take pains to perfect these specific drugs until their application in man also promises a totally specific effect; only now will we ourselves assess the innocuousness of these specific drugs in the clinical department of the Institute for Infectious Diseases. Observations in patients are not needed to establish the specificity of these drugs; they are needed only to confirm it. Next, decisions are required about whether the new drug is already effective enough to elicit unambiguous curative effects in patients. And, when this has been established, we have finally to answer questions of dosage and the most appropriate route of administration. Only when we have accomplished all this do we dare to give our remedy to those doctors who are not yet trained in this kind of drug evaluation for them to test on their own patients . . .'

They provided a detailed account of their approach (Box 1). They presented their results cautiously, emphasizing that all 30 children treated with serum had had diphtheria confirmed bacteriologically and their promising results called for replication on a large-scale:

> So far in the past months 30 cases have been treated with [our] normal serum (or with the equivalent dog curative diphtheria serum of Medical Officer Wernicke): 14 in Berlin, 3 of whom were in Counsellor Henoch's children's department; in all of them the diagnosis 'diphtheria' has been reliably established, in particular, in our Institute, every case in which diphtheria bacilli could not be detected bacteriologically was excluded. Out of these 30 cases, 6 died and 24 have been cured; this is a mortality rate of 20%.
>
> These numbers are still much too small to allow a conclusive judgement about serum therapy: but they are nevertheless encouraging and [they] prompt continuation of serum treatment on an extended scale. Only when we have statistics on hundreds and thousands of diphtheria patients treated with serum will it be the time to deduce final conclusions about the effectiveness of the curative diphtheria serum against this so murderous disease, particularly in childhood.

Historical and concurrent controls to assess the effects of serum treatment for diphtheria

Grancher's Infectious Disease department at the Hôpital des Enfants-Malades in Paris was the site of the first controlled evaluation of the effects of serum treatment for diphtheria.[20] Between February 1 and July 24, 1894 (thus including winter and summer months), Emile Roux, Louis Martin and Auguste Chaillou collected detailed information on 448 children admitted to the diphtheria service. In addition to information about the diphtheritic infection itself, such as duration of the illness, data were collected on age, pulse, breathing rhythm and albuminuria, and information on any complications—from measles, bronchopneumonia, scarlet fever or other co-morbidities. Soon after admission to the hospital, 20 children died but 428 received hyper-immune horse serum in doses ranging from 20 to 125 cc, depending on the severity of the illness and the presence of associated pathologies.

Of the 448 children admitted, 109 died—a fatality of 24.5%. This compared very favourably not only to a rate of about 50% in the same hospital during the 4 years 1891–93, but also to a fatality of 60% in the Hôpital Trousseau, where serum had not been used.

Roux, Martin and Chaillou distinguished diphtheritic sore throat (*angine diphtérique*) from laryngeal diphtheria (croup), the latter being defined by whether or not tracheotomy had been used. They also stressed the different degrees of seriousness, depending on whether the diphtheritic croups were pure or associated with other conditions (in cases associated with staphylococcal and streptococcal infections fatality reached 63% and 80%, respectively). Further analyses of their crude statistics showed that, when consideration was restricted to patients with diphtheritic sore throat, more dramatic differences in favour of the serum emerged—12% died compared to an average of 34% in previous years, and 32% at the Hôpital Trousseau. Figures were also presented for the patients in whom tracheotomy had been used, among whom 49% had died compared to an average of 73% during previous years and 86% at the Hôpital Trousseau.

Roux, Martin and Chaillou further refined their analyses of the 448 children in two ways. First, they identified and removed from their analysis the 128 children in whom there was no bacteriological confirmation of infection with the diphtheria bacillus. Second, they excluded the 20 children who had died soon after arriving at the hospital and who had not received serum. This left 300 patients with bacteriologically confirmed diphtheria who had received serum. These patients experienced a case fatality of 26% compared to ~50% among similar patients in the same hospital over previous years.

Finally, they compared the mortality among 120 children with 'pure' laryngeal diphtheria who had received serum to the mortality among 96 similar children admitted in 1891 and 1892. The case fatality rates were 7.5 and 41%, respectively and the authors provided plausible reasons for the deaths of the nine infants who had died in spite of receiving the serum. Serum treatment had also been associated with a reduced use of tracheotomy.

Unsurprisingly, Emile Roux and his colleagues concluded that this evidence supported their belief that, as serum was the only new element that had been introduced at the Hôpital des Enfants-Malades, the beneficial changes had to be attributed to the treatment.[20] It is worth noting that Roux and his colleagues used the word 'statistiques', albeit without presenting statistical analysis as such.[21]

In September 1894, Emile Roux presented these findings to the International Congress of Hygiene, in Budapest, and this marked the introduction of widespread use of serum therapy in Europe.[20,25,26] In the course of the discussion, which accompanied the report of Roux's lecture, the author mentioned that Hans Aronson of Berlin had reported comparable results concerning the treatment of diphtheria patients with anti-diphtheria serum made in Germany. Aronson mentioned a procedure for obtaining high-titre serum (allegedly three times more efficient than Behring's serum), the use of which had resulted in a decrease in case fatality rate from 40% to 15% among bacteriologically confirmed diphtheritic patients[26–28] (quoting Aronson).

The results obtained in Paris were reflected not only in Berlin, but elsewhere. For example, an American textbook[29] published soon after the French and German results had been reported concluded that the value of anti-toxin serum had been established, but, 'so that readers may themselves to a certain extent have a basis for forming their own opinions', statistics were presented showing trends in fatality among patients admitted to the Willard Parker Hospital for Contagious Diseases in New York, and the Kaiser-und-Kaiserin Friedrich Augusta Hospital in Berlin.[30]

Statistics were frequently used to assess the efficacy of anti-streptococcal and anti-diphtheritic serotherapy in Paris. In contrast, they were rarely used to assess anti-venomous, anti-tetanous and anti-tuberculous serotherapy.[31] In fact, Landouzy refers implicitly to differences in the use of statistics to define treatment effectiveness by referring to the extent to which past experience of the disease provided the basis for reliable inferences about the effects of treatments. In the case of rabies and deadly venom inoculation, the alternative facing physicians was to treat victims with inadequately tested treatments or to watch them die. Unsurprisingly, all patients with either of these two conditions were treated with vaccine or sera, with records only of the numbers of survivors and deaths. The effectiveness of these treatments was deduced from the divergence from expectation of the cumulative ratio of survival to mortality, with a discussion of possible explanations of the failures.[32]

The success of serotherapy in tetanus was sporadic and no statistical analysis was even attempted. Landouzy refers to Marmorek's clinical trials of an anti-serum against streptococci, prepared in a similar way to anti-diphtheria serum.[33] Marmorek, who worked under Roux's supervision, compared the mortality rate among all streptococcal infections in the same hospital ward the year preceding the introduction of the anti-serum (5.12%) and during the year of the trial (3.87%). Moreover, the serum was administered only to patients with severe erysipelas. However, no statistical protocol and no homogeneous cohort of patients were defined. The statistics to which Landouzy refers, thus appear quite primitive compared with Roux's studies of serum treatment of diphtheria. This suggests that the evaluative methods applied by Roux were not in common use at the Institut Pasteur at that time.

Within a year of the report of Roux's observations, there were extensive data comparing the mortality of treated cases with historical control data. In 1895, GC Crandall reported that, having 'recently had access to the Library of the Royal College of Surgeons of England, I gathered as fully as possible, statistics upon the use of the anti-diphtheritic serum.' He assembled these data in what was essentially a systematic review, which included 13 comparisons of treated cases to historical controls (Table 1). Unfortunately, Crandall did not provide references for these reports but in some cases, at least there was consideration of the appropriateness of different potential control data. For example, in the report by Washbourn and his colleagues,[34] concurrent control data from other hospitals—as used by Roux—were given. However, 'on account of the varying standards of diagnosis', Crandall decided not to lay much stress on these data by comparison with the historical control data.

Further studies using observational data and an abandoned attempt to do a controlled trial using alternation

The evidence from Paris and other evidence using historical controls did not convince everyone of the value of anti-diphtheritic serum, however. The debate was complicated both because the disease was undergoing spontaneous fluctuations with decreasing virulence and by claims that the success of serum treatment showed that laboratory research was a more promising approach to tackling diseases associated with poverty than the social reforms for which Virchow and others had been calling.[35] Furthermore, deaths had

Table 1 Statistics upon the use of the anti-diphtheritic serum, in Crandall (1895)

		Number of cases treated with serum	Mortality (%)	Previous mortality (%)
Vierordt	Heidelberg	55	14.6	58.0
Ganghofner	Prague	110	12.7	50.0
Wiederhofer	Vienna	100	25.3	42.8
Kossel	Berlin	350	16.7	34.7
Baginsky (quoted by Virchow)	Berlin	303	3.2	47.8
Sonnenburg	Berlin	107	20.6	27.6
Aronson	Berlin	190	14.0	37.0
Ranke	Munich	85	18.8	48.5
Saltmann	Leipsic	122	18.0	
Risel	Halle	114	8.0	
Roux, Martin and Chaillon	Paris	300	26.0	51.7
Lebreton		258	12.0	
Moizard	Paris	231	14.7	50.0
Washbourn, Goodall, Card and others	London	195	18.6	31.1
White	New York	32	25.0	42.7
Withington	Boston	80	16.0	45.0
Total number of cases		2632		
Average mortality (%)			16.8	
Previous average mortality (%)				42.9
Collective report of other observers in different countries		4022	17.1	

Source: Crandall[89]

been attributed to the anti-toxin, some of which attracted wide publicity.[31,35,36]

Copenhagen was one of the places where doubts about the claims made for serum therapy remained. Sceptics emphasized the unpleasant effects of serum therapy, and these meant that even doctors who were themselves ill with diphtheria rejected the therapy.[37] Søren Thorvald Sørensen, professor at the Blegdamshospitalet, conducted numerous investigations and remained unconvinced of the serum's assumed benefit and concerned about its adverse effects.[38–48]

Sørensen conducted studies of anti-diphtheritic serum at the Blegdamshospitalet, first using German serum from October 1894 to February 1895,[38–42,48] then French serum and Danish serum from March 1895 to March 1896.[43–48]

These studies attempted to evaluate the effects of serum by selecting hospital patients who were as comparable as possible with respect to age and symptoms but who had or had not been treated with serum.[37–48] The results of these comparisons using observational data failed to identify convincing beneficial effects of serum, possibly because patients who were more sick had been selected for the serum therapy.

Sørensen reported that 17 of 51 patients (33%) treated with German serum had died compared with 15 of 46 patients (33%) receiving no serum during the same period. Of the patients who had received French or Danish serum, 9 of 36 (23.8%) had died compared with 5 of 19 patients (26.3%) who had not received the serum during the same period. In both periods, the decision to treat or not was by choice, albeit trying to divide the patients to serum or no serum 'as equally as possible'.[46] Sørensen made clear that the slightly lower estimate of mortality compared with untreated controls in patients who had received French or Danish serum should not be ascribed to the serum used, but rather to the changing character of the epidemic. He appears to have been fully aware of the fallacies of studies based on such observational data for assessing the effects of interventions, mentioning allocation biases and fluctuations in disease severity. Accordingly, he went on to alternate patients to receive or not receive serum:[46]

> During the last months of the experimental phase [with French and Danish serum, likely November 1895 to March 1896] we also tried to select every second severe case for serum, but under the available circumstances this method seemed less

> successful. On the one hand, the method was difficult to carry out, and a subjective factor could not be excluded; on the other hand, we obtained only a few usable cases for our statistics, and the cases seemed far more biased than the ones arbitrarily [Danish: vilkårlige] selected.

Owing to these problems, Sørensen abandoned his attempt to do a controlled trial and did not report separately the number of patients allocated by alternation, nor their clinical results.[46,48] Fibiger later referred to these problems as 'practical difficulties'.[37]

Sørensen's conclusions after these studies were clear and balanced. Although they provided no evidence that serum therapy had had a beneficial effect on either the course of the disease or the risk of death, this absence of evidence could not be taken as evidence that there was no beneficial effect: the experiments had been too few in number; most patients had been selected for serum treatment or no serum treatment using subjective clinical assessment and the number of deaths had been too few to provide reliable statistics.[37,48]

Controlled trials to assess the effects of serum treatment of diphtheria

These uncertainties prompted Johannes Fibiger, professor Sørensen's junior colleague, to propose that further, more rigorously controlled research was needed.[37] Professor Sørensen consented to Fibiger's plan, as long as Fibiger himself carried out the experiment.[37] Hróbjartsson, Gøtzsche and Gluud have reported elsewhere about what ensued.[8]

The introduction to Fibiger's report explains why he had remained unconvinced by the evidence provided by Emile Roux and his colleagues.[37] Fibiger acknowledges that the comparison of serum-treated patients with concurrent patients not-so-treated provided the basis for a potentially dependable verdict on the effects of the serum. However, he was concerned that the introduction of serum treatment at the Hôpital des Enfants-Malades had coincided with improvements in isolation routines and hygiene, so that 'the evidential weight of the experiments was lost'. [Considering that Roux and his colleagues had obtained their non-serum treated controls from another hospital (the Hôpital Trousseau) Fibiger might also have drawn attention to the fact that the Hôpital Trousseau was located in a working-class area of Paris; but he did not.]

Fibiger summarized a number of reports from the USA, Germany, Norway and Denmark suggesting that diphtheria had become less aggressive at the end of the 19th century, as well as referring to Sørensen's unconvincing results.[37] He concluded that 'a new series of experiments had to be planned, and planned in such a way that the result would be absolutely conclusive'.[37]

Fibiger's introduction sets out the rationale for the methodological features of his trial:

> Even with minimal knowledge of diphtheria epidemics, one will recognise that it is necessary to have (1) large numbers, and (2) a long study period. To compensate for the large seasonal variation in mortality, the study should last at least one year. Truly, the control cases in the earlier studies were selected to be as similar as possible to the ones treated with serum, but to eliminate completely the play of chance and the influence of subjective judgement, one had to use a different procedure. The only method that could be used rationally was to treat every other patient with serum and every other patient in the usual way.
>
> In many cases a trustworthy verdict can only be reached when a large number of randomly [Danish: tilfældig] selected patients are treated with the new remedy and, at the same time, an equally large number of randomly [Danish: tilfældig] selected patients are treated as usual.

The choice of Fibiger's allocation method probably reflects the earlier decision to abandon a trial in which Sørensen had planned to allocate patients alternately to receive or not to receive serum.[43] Whatever the nature of the 'practical reasons' may have been for abandoning this plan, Fibiger proposed and Sørensen accepted that all patients admitted on one day would be treated with serum but none would be so treated the next day.[37] As noted by Hróbjartsson and his colleagues, this arrangement left open the possibility of allocation bias, since physicians could favour the admission of the most severely affected patients on the days that serum was being used. Since Fibiger was also aware of the possibility of observer bias in this unblinded trial, he tried to minimize inter-observer variation by using 'concordant observations' by the consultant and himself.

Between May 13, 1896 and May 13, 1897, 1004 patients were admitted to the Blegdamshospitalet with presumed diphtheria. Fibiger excluded 520 of these patients from the analysis, and gives a full account of the reasons. Exclusions were mainly made because the diagnosis had not been confirmed bacteriologically (493 patients), but other patients were excluded because they were moribund on admission or had additional serious infections. The remaining 484 patients—all with bacteriologically confirmed diphtheria and croup—were included in Fibiger's analysis.

These arrangements led to a comparison of well-matched groups of 239 patients who received serum with 245 patients who did not. There were eight deaths in the serum group (3%) and 30 deaths in the control group (12%).[4,37] Using terminology which antedates its more specific meaning today,

Fibiger concluded that 'no objection can be raised against the statistical significance of the numbers', which were deemed correct by an inspector from the Sick Benefit Association.[4,37] However, this beneficial effect came at a cost: at least 145 out of the 239 patients (60%) who had been treated with serum developed serum sickness.

Alternate allocation trials and uncertainty in America

It is notable that Sørensen's and Fibiger's studies were conducted in the setting of scepticism surrounding the use of serum, as such scepticism also led to a similar trial being carried out in America during the latter half of the 1890s. William H Park (1863–1939), who would go on to become the influential director of the laboratories of the New York Board of Health,[49] was by the mid-1890s an early supporter of diphtheria anti-toxin and Diphtheria Diagnostician at the Willard Parker Hospital in New York City. At the same hospital, attending physician Joseph E Winters[50,51] had become a very prominent and vocal opponent of anti-toxin, first engaging with Park in public debate before the New York Academy of Medicine in April 1895. By May of 1896, Winters concluded a long paper read before the New York Academy of Medicine by saying that if he had 'found that antitoxin did not do any harm, even though it was valueless in the treatment of diphtheria – even though it did not reduce the mortality – I would never had said anything against it. It is because I believe it is dangerous that my convictions compel me to speak. The time will come, gentlemen, when every member of this academy will feel with reference to it as I do tonight'.[52]

Describing this background 35 years later, Park would relate his response to Winters' claims:

> An interesting experience developed in the Willard Parker Hospital during the winter of 1896. One of the leading paediatricians of the city and an attending physician at the hospital was violently opposed to the use of antitoxin, and so I arranged that alternate patients should receive the antitoxin, and the remainder should not receive it. The test lasted six weeks; several of the patients who were given the antitoxin early did surprisingly well, while several with similar cases who did not receive antitoxin did badly. The difference in the outcome of the cases was so great that we decided to discontinue the observations. We believed that although we had lost a few lives by it, we had gained a certainty as to the value of antitoxin which we would not otherwise have obtained, and this enabled us to persuade the members of the medical profession much more rapidly than if we had not carried out the experiment.[53]

Yet, the actual sequence of Park's reporting of the study is perhaps instructive regarding the perceived value and acceptance of such alternation of patients in America at the time it was purportedly conducted. Despite contributing frequent reports to the *'Medical Record'* and *'Medical News'* during the latter half of the 1890s, Park never reported the study in those journals; nor did he report on the study in his reviews of the treatment of diphtheria in Alfred L Loomis' and William Gilman Thompson's *System of Practical Medicine*[29] or Hobart Hare's *System of Practical Therapeutics,*[54] despite his inclusion of 'statistics' for the year 1896 in the latter report. Indeed, it seems that the most influential clinical investigation in favour of anti-toxin in America during the latter half of the 1890s would be the 1896 collective investigation by the American Pediatric Society concerning the use of anti-toxin in private practice, gathering and collating case series mortality data from 613 different physicians.[51,55,56]

Park seems to have first publicly described his alternate allocation study at a 'symposium on serum therapy' before the New York County Medical Association in what appears to have been early 1900, where he related:

> A very interesting test of the value of antitoxin in diphtheria . . . was tried a year ago last summer. For 6 weeks only every alternate case received antitoxin. Dr. Winters looked after the treatment of those not receiving it, and Dr. Berg, I believe, those receiving it. I carefully watched both series of cases, and the difference was very marked in favour of the antitoxin series. Even Dr. Winters did not ask to have the test prolonged.[57]

Such a recounting would seem to place the trial in the summer of 1898. Park did not cite Fibiger in his 1900 report, but by the time he delivered his 1906 Harvey Lecture, he related: 'An absolutely ideal method to show the influence of antitoxin is one made by Fibinger [*sic*]. . . . In this, at the same time every other case [*sic*] was treated with antitoxin. . . . This method, however, for obvious reasons is not available at this time. We once made a similar test at the Willard Parker Hospital'.[58] Park cited a 1904 German review paper[59] as his source for the Fibiger information; and by 1912, in a paper justifying the use of animal experimentation, Park reported his own study as having taken place in 1895.[60]

An anonymous report[61] in the 'Echoes and News' section of the April 1, 1899 edition of the *Medical News* on 'The chlorin treatment of diphtheria' related:

> The chlorin treatment of diphtheria has been submitted to a test experiment at the Willard Parker Hospital, New York City, during the past six weeks. Every alternate patient ill with diphtheria

> received the chlorin treatment and every alternate one the antitoxin treatment. In connection with the chlorin treatment tonics and general sustaining remedies were used. Dr. Winters supervised the chlorin treatment, and Dr. Berg the antitoxin. The mortality was higher, and the patients generally seemed to be decidedly worse with the chlorin treatment than those with the antitoxin. It is the opinion of those who witnessed the results that the chlorin solution seemed harmless, but useless. It produced neither good nor bad results. The course of the disease presented the picture so familiar previous to the discovery of antitoxin.

It is not clear whether the 'chlorin' experiment was the original alternate allocation study later recounted by Park, or whether it was a follow-up study. Park may have conducted his original alternate allocation study in 1895, 1896, 1898 or 1899. These multiple cited dates may reflect simple misattribution, or possibly an attempt to establish precedence over Fibiger. Indeed, it appears that the very idea of using alternation of patients may have come from Park's sceptical colleague Joseph Winters.[62] For our purposes here, however, what is most relevant is that Park's apparent reluctance to report the study before 1900 reveals that the calculus concerning the merits and drawbacks (from ethical to logistical) to the methodology—especially when compared with the methodology of 'collective investigation'—was not as unequivocably favourable at the time as it would become in retrospect.

Park would become increasingly familiar and comfortable with alternate allocation studies. In his 1906 Harvey Lecture, he referred to a recent compilation of findings from alternate allocation studies of serum therapy for plague carried out in India, which he may have encountered in the Indian plague commission report,[63] although no citation is given. By the 1920s, Park would play a key role in advancing the use of alternate allocation in the evaluation of anti-pneumococcal anti-serum,[64–66] while also apparently using the method to test the utility of scarlet fever anti-toxin in 1925.[67] His halting reporting of the use of alternation thus lends important nuance to the emergence of what must have seemed a powerful, yet in certain ways, problematic methodology during this time.

A controlled trial to evaluate ways of reducing the adverse reactions to anti-diphtheritic serum

Reactions to serum, some of them fatal,[68] were a significant problem, and ways of reducing them were explored. In 1896, Almroth Wright[69] had proposed that calcium chloride might reduce urticaria and other reactions to sera, and his findings were replicated in Paris in 1904 by Arnold Netter. Netter's doctoral student—Maurice Cousin—conducted a well controlled clinical trial to assess the value of prophylactic calcium chloride in patients admitted to the Diphtheria Pavilion of the Hôpital Trousseau.[70,71] Cousin's doctoral thesis described the procedure adopted (Box 2) and the rationale for and execution of alternate allocation to avoid allocation bias was summarized by Arnold Netter the following year:[71]

> To reassure ourselves of the value of this treatment and to eliminate all risks of error, we used the method of alternation [*la méthode alternante*] ... Only the order of admission of the children determined their assignment to one or to the other category.

Box 2 Extract from Maurice Cousin's doctoral thesis (Source: Cousin[69])

A. METHOD OF ADMINISTRATION OF CALCIUM CHLORIDE. MODE OF COMPARISON

We have seen, in the previous chapter, which eruptions have been very variable, depending on the serum used. They can also depend, to a certain extent, on the severity of the diphtheria. To make sure to eliminate causes of error in comparing results obtained with and without calcium chloride, this is how M. NETTER decided to proceed:

> The children were divided into two categories: *even* numbers and *uneven* numbers. These numbers represented, not the numbers of their beds, but the order in which they were admitted to the Diphtheria Pavilion, as from January 1905.
>
> The twenty first child to be admitted, for example, having this number based on [his] order in the admission register, will be added to the uneven numbers, however serious his disease. The one who is admitted after him, the twenty second therefore, will be placed among the even numbers, regardless of any other consideration.
>
> In this way, two series of children will be obtained: in one as in the other one will find mild cases and severe cases, vigorous children and frail children, etc.
>
> Furthermore, children joining one or other series on the same day are inoculated with the same serum, and this exposed to exactly the same risks. In a word, the conditions are absolutely the same.
>
> These two categories once assembled, the one, the even series, will receive calcium chloride, the other, the uneven series, does not receive it.'

> Of 516 children admitted between 15 January and 31 December 1905 to the diphtheria pavilion at the Trousseau Hospital who received injections of anti-diphtheria serum, 258 thus received calcium chlorate; 258 serving as controls did not receive this treatment.

The Cousin/Netter controlled trial demonstrated that fewer children (12/258) who had been allocated calcium chloride experienced serum reactions than the control children (41/264).

A similar study done at the Wilhelmina Hospital in Amsterdam and referring to the Cousin/Netter controlled trial was reported 2 years later by Dr J Gewin:

> Of the 200 patients who were all treated (injected with) Spronck's anti-diphtheria serum, one half (even numbers) received calcium, the other half (uneven numbers) [did] not [receive it], without considering gender, age or [the severity of] the throat affection. As much as possible the children were given in the same order the serum which was given on that same date [day]; thereby we wanted to achieve that the children of one of the two groups (categories) did not receive per chance a very poisonous serum . . . Approximately half of the children in each age group received calcium chloride, the other half not.[71]

Blinded, controlled comparisons of different anti-diphtheritic sera

With the uptake of serum treatment for diphtheria from the end of the 19th century onwards, deaths from the disease fell dramatically,[3] albeit less dramatically in countries, such as the UK, in which serum treatment had not been adopted wholeheartedly.[35] Nevertheless, although there was wide-spread acceptance that anti-diphtheritic serum was effective, the orthodox explanation of its mechanism of action was challenged by the results of further clinical trials done by Adolph Bingel (1879–1953), head of internal medicine (1910–48) at the City General Hospital in Brunswick (Germany).[73,74]

Bingel was an original and extremely methodical thinker and practitioner. In 1921, he reported on his first 40 cases of pneumo-encephalography, a procedure he had co-invented; and 2 years later he published the results of his first 100 liver biopsies. It is not clear what stimulated his interest in the serum treatment of diphtheria, but this interest endured during the rest of his career.[75]

The accepted account of the serum's mechanism of action was that specific anti-toxins produced by artificially infected animals neutralized toxins released by diphtheria bacteria in infected humans (measured using a method developed by Ehrlich to measure the anti-toxin content of the blood serum). The development and testing of this theory in animals had led to the award of the first Nobel Prize in physiology or medicine, in 1901, to Behring. Acceptance of the theory had been reflected in the establishment of serum factories all over the Western world, using horses (for quantitative reasons) as 'anti-toxin producers'.

Bingel, however, was sceptical of the orthodox view that these specific 'healing-sera' were responsible for the serum's anti-diphtheritic effects.[73] During a severe epidemic in the winter of 1910–11, variable responses to the anti-toxin serum had prompted him to wonder whether its beneficial effect was solely due to the anti-toxin. Might it not be caused by non-specific action resulting simply from administering a serum from another species to patients? Bingel was aware of 'the enormous influences of (foreign) protein from strange (non-human) species', as manifested in serum disease and its marked effects on haematological indices. He was at pains to emphasize that he had not the slightest intention of casting doubts on the results of animal research on immunity, but noted that the variable clinical picture of human diphtheria, 'with its numerous and diverse complications', was completely different from the 'experimentally induced infection or intoxication of an animal', so 'whether a drug influences a human disease can only be decided in man'. This reasoning of species-specific pathophysiology and therapy were the basis for Bingel's decision to undertake a trial comparing (Behring's) 'anti-toxin serum' with 'normal serum', that is, serum derived from horses that had not been infected with diphtheria: 'If no differences are found, the anti-toxin cannot be the effective agent'.[73]

Given the widely acknowledged effectiveness of Behring's anti-toxin serum, Bingel proceeded cautiously, using alternation to create comparable groups of patients:

> After I had treated some adult diphtheria patients with ordinary horse serum in 1911, I began in 1912 to treat alternate adult patients with antitoxin serum and with ordinary serum, exactly in the temporal sequence in which they were admitted to the ward. The children all received antitoxin serum. In the second half of the year 1912 and in the first half of 1913, I gradually lowered the age of those to be treated with ordinary horse serum, and from 1 July 1913 [till 31st December 1916 when he stopped the trial], every second case was treated with ordinary horse serum, whether child or adult, regardless of the severity of the illness or the presence of complications.

Bingel noted that:

> . . . it is absolutely inadmissible to compare the results for different time periods, for example to give antitoxin serum during one year, and then to give only ordinary horse serum during a second year,

> and then to compare the results. That would lead to seriously wrong conclusions, for in no infectious disease is the nature of the epidemic so changeable as in diphtheria. Mostly we see light epidemics, but quite serious ones still occur. I remind [the reader] of the heavy epidemics in Berlin and Hamburg of the year 1910, and the one in Leipzig of 1914, which recall the bad times of the period before serum.[73]

In addition to using alternation to address this problem, Bingel also took steps to reduce observer biases. He noted that it was 'extraordinary difficult . . . to evaluate the influences of therapy on disease unless they are obvious, as for example, the success of a surgical operation or cure of syphilis with mercury or Salvarsan. The therapeutic optimist very easily sees improvement, and the sceptic sees nothing'. In order to reduce these problems, Bingel concealed the identity of the two sera from his assistants and nurses, using the cover names of 'old serum' (for the antitoxin serum) and 'new serum' or 'red serum' (for normal serum).

> To make the trial as objective as possible, I have not relied on my own judgment alone, but have sought the views of the assistant physicians of the diphtheria ward, without informing them about the nature of the serum under test (namely the ordinary horse serum). Their judgement was thus completely without prejudice. I am keen to see my observations checked independently, and most warmly recommend this 'blind' method for the purpose. Even the chief physician may try to draw conclusions about the nature of the serum (unknown to him) that has been used in a particular case: he will be astonished to see how little he is able to do this . . . .Neither I nor my assistants Dr Reusz, Dr Schwab, Dr Weber, Dr Lube could detect a difference between the two sera. Dr Koennecke thought the old (antitoxin) serum had a certain advantage, while Dr Rehder declared that if he were to fall ill, he would wish to be treated with the new (horse) serum. The views of these two gentlemen thus neutralised each other.[73]

Although Bingel did not mention 'blinding' the patients participating in his trial, it is clear from the context that it was a blinded trial in which patients, caregivers and observers were blinded to the intervention, and he recommended that others should use blinding in replications of his study.[73]

Bingel insisted on using measurable criteria 'in order to achieve an objective overall assessment', which he contrasted with 'impressions' from the bedside. His final report was based on an analysis of 471 patients treated with anti-toxin serum and 466 with normal serum.[73] The results were meticulously analysed and presented in detailed tables, as well as in diagrams and illustrative case reports. No marked differences were detected between the impacts of the two sera on the time to shedding of the diphtheritic membranes, or on mortality: there were 47 deaths out of 471 patients (10.0%) given anti-toxin serum, compared with 49 deaths out of 466 patients (10.5%) given normal serum. Nor were differences detected in subgroup analyses in patients who had had tracheotomies or other complications, or after considering the sources of their infections (e.g. from within families). Bingel was also aware of the need to study large numbers of patients to reduce the effects of chance, and claimed that his sample had been 'sufficiently large to prove that no preference can be claimed for anti-toxin serum'.[73]

Bingel's challenging results provoked strong and often emotional reactions, which were reported in the lay press. After all, an achievement crowned with a Nobel Prize, and thus the prestige of German basic research, was at stake. In general, the paper was simply dismissed. In a response,[76] Bingel pointed out that he had never contested either the therapeutic or the prophylactic value of Behring's anti-toxin serum. He acknowledged a valid criticism that the normal serum might have contained some anti-toxin, and admitted that such a possibility had not occurred to him as he assumed that what he had bought from industry labelled as 'normal horse serum' was what had been advertised. To address the criticism, he commissioned analyses of samples of the normal serum he had used. These did indeed reveal a very low concentration of anti-toxin in some of the samples tested—~1–3 international units per cubic centimetre. Bingel deemed this to be so low that it could not have had any material effect when compared with an anti-toxin concentration of 500 international units per cubic centimetre, with patients receiving an average total dose of between 2000 and 8000 international units.[76] This argument prompted some of Bingel's critics to admit that anti-toxin serum must contain other unspecified therapeutic elements besides the anti-toxin, but it did not stop the defenders of anti-toxin serum continuing to dismiss his provocative findings. As Bingel pointed out later,[77] their views implied that 1–3 international units of serum were as effective as 500 U. He was also criticized for withholding from patients a proven effective therapy in order to test a pathophysiological hypothesis. In response, he asserted that the final decision about the therapeutic value in man of a drug stemming from animal experiments, however well justified theoretically, remained with clinicians and that it was also the clinicians' business, and not that of serologists, whether such studies were to be regarded as consistent with medical ethics.[77]

Bingel's defence was clearly persuasive to some clinicians, however, and his request that his trial be

replicated was taken up by some other researchers. Hottinger and Toepfer,[78] for example, reported four separate trials, the largest of which alternated 400 patients to either anti-toxin serum or normal serum. Like Bingel, they were unable to detect any differential effects of the two sera. In the 1940s, Bingel himself performed two further large trials comparing anti-toxin serum and normal serum.[74,75] After he had analysed the cases assembled in his three comparative trials, he concluded that anti-toxin was not the active agent in serum therapy of human diphtheria; the sera acted non-specifically 'as a stimulant activating the defence forces [of the body]'. As in his 1918 trial, Bingel allocated patients in his two later trials to anti-toxin or normal serum using alternation, a method to which he now sought to give credibility by referring to Paul Martini's[78] support of the approach: 'With this ''alternating method'' I believe I would achieve statistically irreproachable [*einwandfrei*] results (see Martini)'.[76]

Adolph Bingel's writings in 1918, and subsequently, reveal considerable methodological and epistemological sophistication. His application, a century ago, of a controlled, double-blind clinical trial involving substantial numbers of patients demonstrates the sophistication developed in the study of therapeutic innovations, years before the advent of formal randomization.

The need for international comparative research on the evolution of methods for testing treatments

In this article, we have used examples selected from the development of treatments for diphtheria to illustrate the development of methods for unbiased assessments of the effects of treatments. Our examples have illustrated the use of case reports and case series to demonstrate dramatic effects of treatment; historical controls and concurrent (non-randomized) controls to assess measures to reduce cross-infection and the effects of anti-toxin serum; alternation to control allocation bias in evaluating the effects of anti-toxin sera, treatments for adverse reactions to serum and the relative merits of different sera; double blinding to control observer biases and recognition of the need for large numbers of observations—all by 1918.

The examples we have cited make clear that at least some researchers in France, Denmark, Germany and the USA were espousing principles and applying methods at the beginning of the 20th century that would eventually lead, during the second half of the century, to the emergence of the controlled clinical trial as we know it today. With some important exceptions, most existing histories of clinical trials have focused on the use of random allocation to generate treatment comparison groups, thus generally ignoring the prior history of alternation to create comparison groups during the first half of the 20th century.[80] The use of alternation 'to avoid the imputation of selection' in generating treatment comparison groups began at least as early as the middle of the 19th century[81] and possibly earlier,[82] and not only in Europe and North America, but also, for example, in the Malay States.[83,84] India was the location of a particularly large number of such alternate allocation studies from the late 19th century onwards, for cholera and plague vaccines and plague serum therapy. The latter is represented by NH Choksy's report[85] reprinted in the current issue of the *IJE*, with the commentaries on it providing details regarding the extraordinary flurry of related activity at this time in India.[86–88] Emerging resources, including full text-searchable journals, are now enabling more extended research into the development of such methods as some among many for promoting 'rational' therapeutics at the end of the 19th and beginning of the 20th centuries.[56,89]

There is also a noticeable lack of documentation and analysis of the evolution of controlled trials in countries other than Britain and the USA.[90–94] We hope that this 'taster' of other material that awaits discovery, documentation and analysis may help to stimulate research into the evolution of controlled clinical trials in countries other than Britain and the USA. As Ilana Löwy[95] has observed, 'transnational comparisons may display unexpected differences and/or surprising similarities; questions initially studied in one context can acquire a different meaning when transposed to another situation; a juxtaposition of developments in several sites can provide information impossible to obtain in single-site studies'.

The kind of international approach we believe is needed is illustrated in Kaptchuk's history of blind assessment and placebo controls;[96] indeed, it was Kaptchuk who drew our attention to Adolf Bingel's remarkable 1918 study. Additional comparative international analyses of the evolution of research methods for testing treatments during the first half of the 20th century would help to provide the background against which a broad international consensus about research methods emerged during the second half of the century.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

Iain Chalmers receives salary support from the English National Institute for Health Research.

Acknowledgements

We wish to thank Estela Dukan at the Sibbald Library at the Royal College of Physicians of Edinburgh for bibliographic assistance; Ted Kaptchuk and Johanne Sanders-Gieschke for information about Adolf Bingel; and Didier Guillemot, Anne Hardy, Ilana Löwy and Esteban Ocaña for helpful comments on earlier drafts of this article.

Conflict of interest: None declared.

References

[1] Tröhler U. *The Introduction of Numerical Methods to Assess The Effects of Medical Interventions During The 18th Century: A Brief History. The James Lind Library, 2010. JLL Bulletin: Commentaries On The History of Treatment Evaluation*. www.jameslindlibrary.org (20 October 2011, date last accessed).

[2] Heiberg P. Studier over den statistiske Undersøgelsesmetode som Hjælpemiddel ved terapeutiske Undersøgelser [Studies on the statistical method of investigation as aid for therapeutic studies]. *Bibliotek Læger* 1897;**89:**1–40.

[3] Russell WT. The epidemiology of diphtheria over the last 40 years. *Medical Research Council Special Report Series No. 247*. London: Stationery Office, 1943.

[4] Hróbjartsson A, Gøtzsche PC, Gluud C. The controlled clinical trial turns 100 years: Fibiger's trial of serum treatment of diphtheria. *BMJ* 1998;**317:**1243–45.

[5] Bretonneau P. Inflammations Spéciales du tissu Muqueux, et en Particulier de la Diphthérite. [Distinctive inflammations of mucous membranes, and in particular concerning diphtheria]. Paris: Crevot, 1826.

[6] Trousseau A. Mémoires sur un cas de trachéotomie pratiquée avec succès dans la période extrême du croup [Report on a case of successful tracheotomy in the latest stages of croup]. *J Connaissances Médico-chirugicales* 1833;**1:**5–9.

[7] Hardy A. Tracheotomy versus intubation: surgical intervention in diphtheria in Europe and the United States, 1825–1930. *Bull Hist Med* 1992;**66:**536–59.

[8] Trousseau A. De la trachéotomie dans la période extrême du croup, et des moyens plus propres à en assurer le succès [About Tracheotomy in The Latest Stages of Croup and About The Best Ways Leading to its Success]. *Archives Générales de Médecine* 1855;**1:**257–68.

[9] Sperati G, Felisati D. Bouchut, O'Dwyer and laryngeal intubation in patients with croup. *Acta Otorhinolaryngol Ital* 2007;**27:**320–23.

[10] Bouchut E. D'une nouvelle méthode de traitement du croup par le tubage du larynx [About a new method of treatment of croup by intubation of the larynx]. *Bull Acad Med* 1858;**23:**1160–62.

[11] Bouchut E. *Leçons cliniques sur les maladies de l'enfance faites à l'hôpital Sainte-Eugénie [Lessons on the Clinics of Children Diseases, given at the Sainte-Eugenie Hospital]*. Paris: Plon, 1858.

[12] Trousseau A. *Du tubage de la glotte et de la trachéotomie, rapport fait à l'Académie impériale de médecine, le 2 novembre [About Intubation of the Glottis and Tracheotomy. Report Made to the Imperial Academy of Medicine on November 2]*. Paris: J-B Baillière et fils, 1858.

[13] Bouchut E. *Traité pratique des maladies des nouveaux-nés, des enfants à la mamelle et de la seconde enfance [Practical Treatise of the Diseases of New-borns, of Children Fed at the Breast, and of Older Children]*. Paris: JB Baillière et fils, 1867.

[14] O'Dwyer J. Fifty cases of croup in private practice treated by intubation of the larynx with description of the method. *Med Rec* 1887;**32:**557–61.

[15] Bourges H. *La diphtérie.Bibliothèque médicale Charcot Debove [Diphtheria. Medical Library Charcot-Debove]*. Paris: Charcot-Debove, 1892.

[16] Sevestre L, Martin L. *Diphtérie: Traité des maladies de l'enfance, tome 1 [Diphtheria: Treatise of Children Diseases]*. Paris: Masson et fils, 1896.

[17] Roux E, Yersin A. Contribution à l'étude de la diphtérie, 3e mémoire [Contribution to the study of diphtheria, 3rd paper]. *Annales de l'Institut Pasteur* 1890;**7:**384–426.

[18] Opinel A. The anti-diphtheria apparatus at the Institut Pasteur, 1894–1905. *Dynamis* 2007;**27:**83–106.

[19] Gachelin G. The designing of anti-diphtheria serotherapy at the Institut Pasteur (1888–1900): the role of a supranational network of microbiologists. *Dynamis* 2007;**27:** 45–62.

[20] Roux E, Martin L, Chaillou A. Trois cent cas de diphthérie traité par le serum antidiphthérique [Three hundred cases of diphtheria treated using anti-diphtheria serum]. *Annales de l'Institut Pasteur* 1894;**8:**640–61.

[21] Opinel A, Gachelin G. *French 19th Century Contributions to the Development of Treatments for Diphtheria. JLL Bulletin: Commentaries on the History of Treatment Evaluation* 2010; (http://www.jameslindlibrary.org).

[22] Grancher J. Essai d'antisepsie médicale [A trial of medical antisepsis]. *Revue d'hygiène et de police sanitaire* 1890;**12:** 495–509, (and commentaries pp. 992–1003).

[23] Grancher J. Prophylaxie de la diphtérie. Transport et isolement des diphtériques dans les hôpitaux [Prophylaxis of diphtheria. Transportation and isolation of diphteritic patients in hospitals]. *Revue d'hygiène et de police sanitaire* 1890;**12:**1084–97.

[24] Behring EA, Boer, Kossel H. Zur Behandlung diphtheriekranker Menschen mit Diphtherieheilserum. *Deutsche Medicinische Wochenschrift* 1893;**17:**389–93.

[25] Behring EA. In anonymous, discussion. *Revue d'hygiène et de police sanitaire* 1894;**16:**784–98.

[26] Anonymous. Echoes and news: the chlorine treatment of diphtheria. *Med News* 1899;**74:**403.

[27] Hüntelmann A. Diphtheria serum and serotherapy. Development, production and regulation in *fin de siècle* Germany. *Dynamis* 2007;**27:**107–31.

[28] Gachelin G. The construction of a culture of standardization at the Institut Pasteur 1885-1900. In: Gradmann C., Simon J. (eds). *Evaluating and standardizing therapeutic agents, 1890–1950*. Basingstoke, UK: Palgrave Macmillan, 2010, pp. 71–88.

[29] Loomis AL, Thompson WG. *A System of Practical Medicine by American Authors*. New York and Philadelphia: Lea Brothers, 1897, pp. 695–98.

[30] Baginski A. The serum therapy in diphtheria; (cited in Loomis and Thompson, Berlin 1897), 1895.

[31] Landouzy LTJ. *Les sérothérapies: leçons de Thérapeutique et Matière Médicale [Serotherapies : Lessons of Therapeutics and Clinical Matter]*. Paris: Carré et Naud, 1898.

[32] Gamaleya N. Sur les prétendues statistiques de la rage [About the so-called statistics of rabies]. *Annales de l'Institut Pasteur* 1887;**1:**289.

[33] Marmorek A. Le streptocoque et le sérum antistreptococcique. *Annales de l'Institut Pasteur* 1895;**9:**38.
[34] Washbourn JW, Goodall EW, Card AH. A series of 80 cases treated with diphtheria anti-toxin, with observations on their bacteriology. *Br Med J* 1894;**2:**1417–19.
[35] Weindling P. From medical research to clinical practice: serum therapy for diphtheria in the 1890s. In: Pickstone JV (ed.). *Medical Innovations in Historical Perspective*. New York: St Martin's Press, 1992, pp. 72–83.
[36] Hüntelmann A. Das Diphtherie-Serum und der Fall Langerhaus. *Med Ges Gesch* 2005;**24:**71–104.
[37] Fibiger J. Om serumbehandling af difteri [On treatment of diphtheria with serum]. *Hospitalstidende* 1898;**6:**309–25.
[38] Sørensen ST. Forsøg med Serumbehandling (tysk Serum) paa Blegdamshospitalet fra den 16. Oktbr. 94 til 6 Febr. 95. [Experiments with serum treatment (German serum) at Blegdamshospitalet from 16 October 94 to 6 February 95]. *Hospitalstidende* 1895;**4:**293–309.
[39] Sørensen ST. Forsøg med Serumbehandling (tysk Serum) paa Blegdamshospitalet fra den 16. Oktbr. 94 til 6 Febr. 95. [Experiments with serum treatment (German serum) at Blegdamshospitalet from 16 October 94 to 6 February 95]. *Hospitalstidende* 1895;**4:**321–39.
[40] Sørensen ST. Forsøg med Serumbehandling (tysk Serum) paa Blegdamshospitalet fra den 16. Oktbr. 94 til 6 Febr. 95. [Experiments with serum treatment (German serum) at Blegdamshospitalet from 16 October 94 to 6 February 95]. *Hospitalstidende* 1895;**4:**353–62.
[41] Sørensen ST. Forsøg med Serumbehandling (tysk Serum) paa Blegdamshospitalet fra den 16. Oktbr. 94 til 6 Febr. 95. [Experiments with serum treatment (German serum) at Blegdamshospitalet from 16 October 94 to 6 February 95]. *Hospitalstidende* 1895;**4:**378–89.
[42] Sørensen ST. Forsøg med Serumterapi ved Difteritis. Statistik fra Tyskeserumperioden.[Experiments with serum treatment of diphtheria infection. Statistics from the period of German serum]. *Hospitalstidende* 1896;**4:** 73–87.
[43] Sørensen ST. Forsøg med Serumterapi ved Difteritis. Statistik fra Tyskeserumperioden.[Experiments with serum treatment of diphtheria infection. Statistics from the period of German serum]. *Hospitalstidende* 1896;**4:** 521–41.
[44] Sørensen ST. Forsøg med Serumterapi ved Difteritis. Statistik fra Tyskeserumperioden.[Experiments with serum treatment of diphtheria infection. Statistics from the period of German serum]. *Hospitalstidende* 1896;**4:** 543–58.
[45] Sørensen ST. Forsøg med Serumterapi ved Difteritis. Statistik fra Tyskeserumperioden.[Experiments with serum treatment of diphtheria infection. Statistics from the period of German serum]. *Hospitalstidende* 1896;**4:** 578–90.
[46] Sørensen ST. Forsøg med Serumterapi ved Difteritis. Statistik fra Tyskeserumperioden.[Experiments with serum treatment of diphtheria infection. Statistics from the period of German serum]. *Hospitalstidende* 1896;**4:** 602–12.
[47] Sørensen ST. Forsøg med Serumterapi ved Difteritis. Statistik fra Tyskeserumperioden.[Experiments with serum treatment of diphtheria infection. Statistics from the period of German serum]. *Hospitalstidende* 1896;**4:**621–28.
[48] Sørensen ST. Versuche mit Serumtherapie bei Diphtherie im Blegdamspitale[Blegdamshospitalet] in Kopenhagen. [Experiments with serum treatment of diphtheria infection in Blegdamshospitalet in Copenhagen]. *Therap Monatshefte März* 1896;**10:**119–25.
[49] Oliver WW. *The Man Who Lived For Tomorrow: A Biography of William Hallock Park*. New York: E.P. Dutton and Company, 1942.
[50] Winters JE. Discussion on the use of diphtheria antitoxin. *Trans New York Acad Med* 1895;**12:**204–15.
[51] Hammonds EM. Childhood's deadly scourge. *The Campaign to Control Diphtheria in New York City, 1880–1930*. Baltimore: Johns Hopkins University Press, 1999.
[52] Winters JE. Clinical observations upon the use of antitoxin in diphtheria: A report of personal investigation of this treatment in the principal fever hospitals of Europe during the summer of 1895. *Med Rec* 1896;**49:** 877–93.
[53] Park WH. The history of diphtheria in New York City. *Am J Dis Child* 1931;**42:**1431–45.
[54] Hare H. *System of Practical Therapeutics*. Philadelphia: Lea Brothers and Company, 1897.
[55] American Pediatric Society. Report of the American Pediatric Society Collective Investigation into the use of antitoxin in the treatment of diphtheria in practice. *Arch Pediat* 1896;**13:**481–503.
[56] Marks HM. ''Until the Sun of Science . . . the true Apollo of Medicine has risen'': Collective investigation in Britain and America, 1880–1910. *Med Hist* 2006;**50:**147–66.
[57] Park WH. Use of diphtheria antitoxin in the treatment and prevention of diphtheria. *J Am Med Assoc* 1900;**34:** 902–04.
[58] Park WH. *A Critical Study of the Results of Serum Therapy in Diseases of Man. Harvey Lectures, 1905–1906*. Philadelphia: JB Lippincott and Company, 1906.
[59] Faber EE. Der Einfluss der Serumbehandlung auf die Diphtheriemortalität [The influence of serum treatment on the mortality of diphtheria]. *Jahrbuch für Kinderheilkunde und Physische Erziehung* 1904;**59:**620–35.
[60] Park WH. The role of animal experimentation in the discoveries leading to our present knowledge of the etiology, prevention and cure of diphtheria. *J Am Med Assoc* 1912; **58:**453–59.
[61] Anonymous. Echoes and news: the chlorine treatment of diphtheria. *Med News* 1899;**74:**403.
[62] Podolsky SH, Davey Smith G. Park's story and Winters' tale: alternate allocation clinical trials in turn of the century America; The James Lind Library, 2010. JLL Bulletin: Commentaries On The History of Treatment Evaluation. www.jameslindlibrary.org (20 October 2011, date last accessed).
[63] Indian Plague Commission. *Minutes of Evidence Taken by the Indian Plague Commission with Appendices*. London: Eyre and Spottiswoode, 1900–1901.
[64] Park WH, Bullowa JGM, Rosenblüth NM. The treatment of lobar pneumonia with refined specific antibacterial serum. *J Am Med Assoc* 1928;**91:**1503–08.
[65] Podolsky SH. *Pneumonia Before Antibiotics: Therapeutic Evolution and Evaluation in Twentieth-Century America*. Baltimore: Johns Hopkins University Press, 2006.
[66] Podolsky SH. *Jesse Bullowa, specific treatment for pneumonia, and the development of the controlled clinical trial*; *The James Lind Library, 2008 JLL Bulletin: Commentaries On The History of Treatment Evaluation.* (www.jameslindlibrary.org) (20 October 2011, date last accessed).
[67] Park WH. Scarlet fever: etiology, prevention by immunization, and antitoxic treatment. *J Am Med Assoc* 1925;**85:** 1180–86.

[68] Weindling P. From medical research to clinical practice: serum therapy for diphtheria in the 1890s. In: Pickstone JV (ed.). *Medical Innovations in Historical Perspective*. New York: St Martin's Press, 1992, pp. 72–83.
[69] Wright AE. On the treatment of the haemorrhages and urticarias which are associated with deficient blood coagulability. *Lancet* 1896;**1:**153–55.
[70] Cousin M. *Des éruptions consecutives aux injections de sérum antidiphthérique et de leur traitement prophylactique par l'ingestion de chlorure de calcium. Thèse pour le Doctorat en médecine*. Paris: Jules Rousset, 1905, pp. 36–44.
[71] Netter A. Efficacité de l'ingestion de chlorure de calcium comme moyen préventif des éruptions consecutives aux injections de sérum. *Séances et Mémoires de la Société de Biologie* 1906;**58:**279–80.
[72] Gewin J. Chlorkalzium gegen die Serumkrankheit. *Münchner Medizinische Wochenschrift* 1908;**50:**2670–71.
[73] Bingel A. Über Behandlung der Diphtherie mit gewöhnlichem Pferdeserum. *Deutsches Archiv für klinische Medizin* 1918;**125:**284–332.
[74] Tröhler U. *Adolf Bingel. James Lind Library, 2010. JLL Bulletin: Commentaries On The History of Treatment Evaluation*. www.jameslindlibrary.org (20 October 2011, date last accessed).
[75] Tröhler U. *Adolf Bingel (1879–1953). Adolf Bingel's Blinded, Controlled Comparison of Different Anti-diphtheritic Sera in 1918. JLL Bulletin: Commentaries On The History of Treatment Evaluation*. www.jameslindlibrary.org (20 October 2011, date last accessed).
[76] Bingel A. Zur Behandlung der Diphtherie mit gewöhnlichem Pferdeserum. *Deutsche Medizinische Wochenschrift* 1919;**45:**739–40.
[77] Bingel A. Wirkt das Diphtherieheilserum bei der menschlichen Diphtheriekrankheit spezifisch durch seinen Antitoxingehalt oder unspezifisch? *Deutsche Medizinische Wochenschrift* 1949;**79:**101–03.
[78] Hottinger A, Toepfer D. Über den Wert der Serumtherapie bei Diphtherie, insbesondere bei der malignen, toxischen Form. *Zeitschrifte für Kinderheilkunde* 1933;**54:** 505–29.
[79] Martini P. *Methodenlehre der Therapeutischen Untersuchung*. Berlin: Springer, 1932.
[80] Chalmers I. *Why the 1948 MRC trial of streptomycin for pulmonary tuberculosis used treatment allocation based on random numbers; The James Lind Library, 2010 JLL Bulletin: Commentaries On The History of Treatment Evaluation* (www.jameslindlibrary.org) (20 October 2011, date last accessed).
[81] Balfour TG. *Quoted in West C. Lectures on the Diseases of Infancy and Childhood*. London: Longman, Brown, Green and Longmans, 1854, p. 600.
[82] Milne I, Chalmers I. *Hamilton's Report of A Controlled Trial of Bloodletting, 1816. The James Lind Library, 2002. JLL Bulletin: Commentaries On The History of Treatment Evaluation*. www.jameslindlibrary.org (20 October 2011, date last accessed).
[83] Fletcher W. Rice and beri–beri: preliminary report on an experiment conducted in the Kuala Lumpur Insane Asylum. *Lancet* 1907;**1:**1776–79.
[84] Vandenbroucke JP. *The contribution of William Fletcher's 1907 report to finding a cause and cure for beriberi*. The *James Lind Library, 2003 JLL Bulletin: Commentaries On The History of Treatment Evaluation*. www.jameslindlibrary.org (20 October 2011, date last accessed).
[85] Choksy NH. Letter to the Editor: Professor Lustig's plague serum. *Lancet* 1900;**156:**291–92. Reprinted *Int J Epidemiol* 2013;**42**:654–56.
[86] Chakrabarti P. Commentary: an experimental theatre for vaccines: Bombay in the time of plague. *Int J Epidemiol* 2013;**42:**656–58.
[87] Ramanna M. Commentary: NH Choksy and serum therapy. *Int J Epidemiol* 2013;**42:**658–59.
[88] Syed IS, Swaminathan KL. Commentary: Dr Choksy's dilemma. *Int J Epidemiol* 2013;**42:**659–62.
[89] Marks HM. *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*. Baltimore: Johns Hopkins University Press, 1997.
[90] Marks HM. Ideas as reforms. Therapeutic experiments and medical practice, 1900–1980. PhD Thesis. Massachusetts Institute of Technology, 1987.
[91] Cox-Maximov D. The making of the clinical trial in Britain, 1910–1945: expertise, the state and the public. PhD Thesis. University of Cambridge, 1997.
[92] Toth B. Clinical trials in British medicine 1848–1948, with special reference to the development of the randomised controlled trial. PhD Thesis. University of Bristol, 1998.
[93] Edwards M. Control and the therapeutic trial, 1918–1948. MD Thesis. University College London, 2004.
[94] Williams KJ. British pharmaceutical industry, synthetic drug manufacture and the clinical testing of novel drugs 1895–1939. PhD Thesis. University of Manchester, 2005.
[95] Löwy I. The social history of medicine: beyond the local. *Soc Hist Med* 2007;**20:**481–565.
[96] Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med* 1998;**72:**389–433.
[97] Crandall GC. Serum-therapy – statistics upon the use of the diphtheria antitoxin. *J Am Med Assoc* 1895;**25:** 145–46.