REVIEW

Bell’s palsy: aetiology, clinical features and multidisciplinary care

Timothy J Eviston,1 Glen R Croxson,2 Peter G E Kennedy,3 Tessa Hadlock,4 Arun V Krishnan1

ABSTRACT

Bell’s palsy is a common cranial neuropathy causing acute unilateral lower motor neuron facial paralysis. Immune, infective and ischaemic mechanisms are all potential contributors to the development of Bell’s palsy, but the precise cause remains unclear. Advancements in the understanding of intra-axonal signal molecules and the molecular mechanisms underpinning Wallerian degeneration may further delineate its pathogenesis along with in vitro studies of virus–axon interactions. Recently published guidelines for the acute treatment of Bell’s palsy advocate for steroid monotherapy, although controversy exists over whether combined corticosteroids and antivirals may possibly have a beneficial role in select cases of severe Bell’s palsy. For those with longstanding sequelae from incomplete recovery, aesthetic, functional (nasal patency, eye closure, speech and swallowing) and psychological considerations need to be addressed by the treating team. Increasingly, multidisciplinary collaboration between interested clinicians from a wide variety of subspecialties has proven effective. A patient centred approach utilising physiotherapy, targeted botulinum toxin injection and selective surgical intervention has reduced the burden of long-term disability in facial palsy.

INTRODUCTION

Bell’s palsy is an acute-onset peripheral facial neuropathy and is the most common cause of lower motor neuron facial palsy.1 The clinical presentation of the disorder is a rapid onset, unilateral, lower motor neuron-type weakness with accompanying symptoms of postauricular pain, dysgeusia, subjective change in facial sensation and hyperacusis. This clinical presentation can be explained by the anatomical construct of the human facial nerve, specifically its mixed nerve profile containing motor, sensory and parasympathetic fibres. The propensity for the facial nerve to form numerous connections with adjacent cranial nerves2 may also explain the occasionally observed features of altered facial sensation (cranial nerve V), vestibular dysfunction (cranial nerve VIII) or pharyngeal symptoms (cranial nerves IX and X).3 Reduced lacrimation and salivation secondary to parasympathetic effects may also occur.4 Maximal disability occurs within the first 48–72 h and the severity of the palsy correlates with the duration of facial dysfunction, the extent of facial recovery and impairment of quality of life.

Despite extensive study of the condition, the exact pathogenesis of Bell’s palsy is still controversial.5 Infection (herpes simplex type-1),4 nerve compression6 and autoimmunity7 may all play a role, yet the exact sequence and magnitude of these influences remains unclear.

Anatomy

The human facial nerve is the seventh cranial nerve (CNVII) and comprises motor, sensory and parasympathetic components. Its function is responsible for voluntary and mimetic facial movement, taste to the anterior two-thirds of the tongue, and control of salivary gland and lacrimal gland secretions.

The facial nerve receives axons from the superior part of the solitary nucleus and superior salivary nucleus that form the nucleus intermedio-lateralis compendium (sensory and parasympathetic axons) and motor efferent fibres from the facial nucleus, which receives synaptic input from the contralateral motor cortex for all facial movements except the forehead, which has bicortical input.

The path of the facial nerve has intracranial, intratemporal and extratemporal components. Its intracranial course runs from the pontomedullary angle to the internal acoustic meatus where it is accompanied by the vestibulocochlear nerve (CNVIII). The intratemporal course of the facial nerve is long and tortuous. During its intratemporal course, the nerve encounters the geniculate ganglion and gives rise to the superior petrosal nerve, the nerve to stapedius and chorda tympani nerve branches, before exiting the skull base through the stylorhinal foramen. The extratemporal facial nerve courses through the substance of parotid gland dividing it into deep and superficial lobes. It gives off the posterior auricular nerve and nerve to the posterior belly of digastric before dividing into its terminal facial branches. There is significant variation in the branching pattern of the terminal facial branches, which are traditionally conceptualised into temporal, zygomatic, buccal, marginal mandibular and cervical branches. These terminal motor branches are responsible for all facial expression and functional tasks such as eye and mouth closure and nasal patency during inspiration. Throughout its course, the facial nerve forms multiple communications between its own branches and with adjacent cranial nerves.2

HISTORY

Sir Charles Bell (1774–1842) was fascinated by the nervous system. As an accomplished anatomist, artist, surgeon and teacher, his work on the characterisation of the peripheral nervous system through anatomical study, vivisection and clinical
correlation, provided a significant contribution to medical knowledge of his time. In particular, he was fascinated with the separation of sensory (trigeminal nerve) and motor supply (facial nerve) to the face. It was his eloquent and logical descriptions that elevated his status among his contemporaries and ushered the ‘post-Bell’s’ era, which saw a rapid increase in the number of publications relating to acute idiopathic facial palsy—Bell’s palsy.

Although Bell’s descriptions were admirable in their ability to inspire his contemporaries to document and study the disorder, myriad other detailed descriptions exist in Greek, Persian and European medical texts as far back as the fifth century BCE, and there is prehistoric ceramic art depicting facial palsy identified in ancient Peruvian culture. Other clinicians who recognised the entity of acute idiopathic facial paralysis before Bell (and perhaps may have inspired some of his observations) include Sydenham, Stalpart van der Wiel, Douglas, Friedrich and Thomasen á Thuessink.

Approximately 1000 years before Bell, the Persian physician and scholar, Razi (865–925 CE), described facial palsy at length in his seminal ninth century text ‘al-Hawi’. This remarkable description referenced the contributions of Galen and Celsus, among others, and included a diagnostic algorithm delineating peripheral facial palsy from more sinister central causes, which were accompanied by delirium, coma, hemiplegia, blindness, or deafness, and tended to have a poor prognosis.

**Epidemiology**

Bell’s palsy is a common cranial mononeuropathy. It affects males and females equally, and has a slightly higher incidence in mid and later life, but certainly occurs across all age ranges. The described population incidence rates range from 11.5 to 40.2/100,000 with specific studies demonstrating similar annual incidences between the UK (20.2/100,000), Japan (30/100,000) and the USA 25–30/100,000. Clustering and epidemic phenomena are not demonstrated in the majority of studies. An exception to this is the recent occurrence of an increase in incidence of Bell’s palsy during a trial of intranasal vaccine delivery. This was possibly due to the immune effects of the detoxified Escherichia coli heat labile toxin adjuvant used in this form of vaccine delivery. The incidence is higher in pregnancy, following viral upper respiratory tract infection, in the immunocompromised setting, and with diabetes mellitus and hypertension. There is no distinct latitudinal variation for incidence, nor is there a racial or ethnic predilection. Some epidemiological data demonstrate seasonal variation, with a slightly higher incidence in cold months versus warm months, and a slight preponderance to arid over non-arid climates.

**Aetiology**

There is a diverse body of evidence implicating immune, infective and ischaemic mechanisms as potential contributors to the development of Bell’s palsy, but the cause of classical Bell’s palsy remains unclear. One possible cause that has been suggested is that of a reactivated herpes simplex virus (HSV-1) infection centred around the geniculate ganglion, a theory first outlined by McCormick. The association with HSV-1 is supported by the presence of HSV-1 in intratemporal facial nerve endoneural fluid harvested during nerve decompression, and the ability to incite facial palsy in an animal model through primary infection and reactivation induced by immune modulation. Despite this evidence, the behaviour of Bell’s palsy is unusual compared with other diseases more commonly caused by HSV such as herpes labialis (cold sores) and genital herpes. Further, it is not justifiable to assume a cause and effect relationship between the presence of HSV-1 in the patients’ endoneurial fluid and the development of Bell’s palsy.

HSV-1 is one of several human herpes viruses known to have a neurotrophic capacity for peripheral nerves, and other viruses in this category include herpes simplex virus type 2 (HSV-2) and varicella zoster virus (VZV). They enter the body through mucocutaneous exposure and establish their presence in latent form with highly restricted gene transcription in multiple ganglia throughout the neuroaxis for the entire life of the host, including in the cranial, dorsal root and autonomic ganglia. This latent presence in ganglia in the absence of active viral replication and assembly is characteristic, well described, and widely distributed through normal and diseased populations. HSV has a global distribution and is a fundamentally resilient virus. HSV and VZV can both reactivate in an immunocompetent host, and in the presence of circulating antibodies, although reactivation is more likely in the setting of immunodeficiency, especially in the case of VZV.

A possible cause of neural dysfunction due to HSV-1 is the activation of intra-axonal degradation and apoptotic pathways driven by local direct and indirect responses of the axon to the virus itself in a susceptible phenotype. Although not previously associated specifically with the pathogenesis of Bell’s palsy, the emergence of literature pertaining to the role of intra-axonal signal molecules (eg, SARM1), mitochondrial permeabilisation and the molecular mechanisms underpinning Wallerian degeneration, suggests that acute axonal degeneration in the context of viral infection may be an evolutionarily conserved innate immune response to prevent virus transport to the central nervous system.

Recent in vitro work has demonstrated local messenger RNA transcription in peripheral nerve axons incited by the presence of α-herpes virus particles. In this compartmentalised model, protein and signal transduction changes were not reliant on nuclear machinery, that is, when a virus enters an axon, the axon responds locally. Earlier work examining cellular physiology in the setting of herpes infection demonstrated an acute decrease in sodium conductivity in the setting of HSV-1.

Changes in sodium conductance can result in a reversed sodium–calcium exchange (NCX) current and the accumulation of intracellular calcium. This aberration in calcium homoeostasis leads to protease activation and intra-axonal degeneration. These processes of axonal degeneration would drive the abrupt onset of Bell’s palsy and explain the lack of a pronounced immune response. This would not necessarily discount the role of compression to the pathogenesis but, rather, may answer the question as to why the nerve swells, leading to impingement, in the first place.

Another possible contributor to the pathogenesis of Bell’s palsy implicates the role of a cell-mediated immune response against myelin, akin to a mononeuropathic form of Guillain-Barré syndrome (GBS). The evidence for this stems from the indirect laboratory finding of GBS, such as changes in peripheral blood percentages of T and B lymphocytes, elevated chemokine concentrations and in vitro reactivity to myelin protein (P1L) in blood samples taken from patients with Bell’s palsy.

**Diagnosis**

Bell’s palsy is a clinical diagnosis. The characteristic findings are acute onset of unilateral lower motor neuron facial paralysis that affects muscles of the upper as well as lower face and reaches its peak by 72 h. These findings are frequently accompanied by symptoms of neck, mastoid or ear pain, dysgeusia, hyperacusis
or altered facial sensation. These associated symptoms are present in 50–60% and are reassuring for the diagnosis of Bell’s palsy.

The involvement of posterior auricular, petrosal, chorda tympani and stapedius nerves implicates the site of dysfunction being within the temporal bone. Localisation to the intratemporal facial nerve is further supported by unilateral enhancement of the geniculate, labyrinthine and meatal segments of the facial nerve on contrast-enhanced MRI studies. This imaging finding is hypothesised to represent disruption of the blood–brain barrier and vascular congestion of the facial nerve. Attempts to determine surrogate diagnosis through PCR techniques with VZV and HSV primers applied to posterior auricular, tear or facial muscle specimens have failed to demonstrate any consistent correlation between viral load and clinical features. These tests are therefore of limited value as diagnostic tools.

Differential diagnosis
The differential diagnosis for facial palsy is broad, and misdiagnosis is not uncommon. Causes of facial palsy may be divided into congenital and acquired aetiologies. Congenital causes include genetic syndromes, birth-related trauma and isolated disorders of development (eg, developmental hypoplasia of facial muscles). Acquired causes include infective (VZV, Lyme disease, mycobacterium tuberculosis, HIV), traumatic (iatrogenic or head trauma), inflammatory (vasculitis, sarcoidosis, autoimmune disease), neoplastic (benign or malignant) and cerebrovascular causes, among others. In the experience of one of the authors (GC) in an expert referral setting, the rate of misdiagnosis of Bell’s palsy by the initial consulting clinician is 10.8%. Missed diagnoses include tumours (eg, facial nerve schwannoma, parotid malignancy and, rarely, acoustic neuroma), herpes zoster oticus and granulomatous diseases such as sarcoidosis and granulomatosis with polyangiitis (Wegener’s granulomatosis).

A structured clinical approach that considers the pattern of facial palsy (figure 1) along with patient characteristics and a thorough physical examination will generally provide evidence for an alternative diagnosis, and prompt appropriate investigation. Particular patterns of facial palsy that require thoughtful consideration include: (1) fluctuant, step-wise or slowly progressive (beyond 72 h), facial palsy; (2) bilateral palsy (GBS, carcinomatosis, lymphoma); (3) recurrent facial palsy (facial nerve neuroma); (4) prolonged complete palsy (>4 months) and (5) sudden complete facial palsy (haemorrhage into a tumour). These patterns should prompt a detailed search for an underlying cause. Likewise, the presence of a mass in the parotid region, a history of cutaneous malignancy or segmental facial nerve weakness should raise suspicion for a tumour. A history of trauma, ear symptoms such as ipsilateral deafness, tinnitus, fullness or discharge, or systemic symptoms such as fever, are also red flags warranting further investigation and specialist otorhinolaryngological consultation.

The consideration of cerebrovascular disease as a cause of facial palsy is important with this being the main concern for many patients and clinicians, often prompting expert neurology consultation. The preservation of upper facial movement (frontalis contraction) is a discriminator between cortical (central) and peripheral facial nerve weakness. Rarely, ipsilateral pontine pathology may result in lower motor neuron pattern facial weakness due to direct compromise of the facial nucleus. This will be accompanied by other cranial nerve, and long tract symptoms and signs. Ipsilateral abducens nerve (CNVI) dysfunction (lateral gaze palsy) is a particularly useful sign.

Grading
The most widely used systems for recording the severity of Bell’s palsy are the House-Brackmann (HB) scale or the Facial Nerve Grading Scale (also known as the Sunnybrook system). The subjective nature of these scales makes them prone to some misinterpretation and interobserver variability; however, their ease of use has cemented their role in clinical practice for communicating the overall degree of dysfunction, for monitoring outcomes and for the presentation of group data in a research...
or audit setting. In a recent survey of facial nerve specialists, photography and videography were ubiquitous among respondents and, indeed, the use and importance of video recording of standardised facial movements (eyebrow raise, gentle eye closure, tight eye closure, snarl, open and closed lip smiles, lip depression and lip puckering) is emphasised for the accurate outcome assessment of interventions such as chemodenervation, physiotherapy or surgery, which may be considered in those patients who have an incomplete recovery.

Neurophysiology

There have been a number of studies exploring the potential utility of neurophysiological assessment for treatment selection and prognosis. In the past, nerve conduction studies of the facial nerve, also referred to as electroneuronography (ENoG) in the surgical literature, have been suggested in some studies to be useful in the selection of patients who may require surgical decompression of the facial nerve. The demonstration of a greater than 90% reduction in compound muscle action potential in the first 10 days of onset compared with the unaffected side is associated with a 50% chance of incomplete recovery and was a trigger for operative intervention in some centres.

In contemporary practice, facial nerve decompression has increasingly fallen out of the normal domain due to cost, risk and lack of efficacy. With this trend away from surgery, the time and cost of neurophysiological assessment outweighs the benefit for the vast majority of patients. An exception to this is the clinical setting of complete paralysis. Here, neurophysiology provides useful information with the presence of a residual response on neurophysiology suggesting a predominantly neuropraxic injury, in which case the prospect of a good recovery (HB I or II) is high. The absence of a neurophysiological response is suggestive of complete degeneration and a prolonged paralysis with incomplete recovery that may be complicated by synkinesis. The knowledge of a prolonged recovery may sway the clinician and patient towards surgical eye protection procedures and the proactive involvement of a facial therapist early in the course of recovery.

TREATMENT

Acute treatment

The American Academy of Neurology (AAN) and the American Academy of Otolaryngology—Head and Neck Surgery Foundation (AAO-HNSF) have recently published guidelines for the treatment of Bell’s palsy. Although the structure and scope of these guidelines are different, they are essentially complementary documents that reinforce the role of corticosteroids in the treatment of Bell’s palsy and argue against the routine use of antiviral therapy. Furthermore, the AAO-HNSF guidelines discourage routine laboratory, imaging or neurophysiological testing at the first presentation of typical Bell’s palsy. The dose of oral steroids should be started in the first 72 h of onset and a regime mirroring either of the Scottish or European randomised controlled trials (RCTs) should be used. This is either 50 mg prednisone for 10 days or 60 mg for the first 5 days, then reducing by 10 mg each day for the next 5 days. Both seem effective. It has been argued that the lack of significance demonstrated by combined corticosteroid and antiviral therapy over corticosteroids alone in double-blind RCTs represents a dilution effect of mild and moderate palsies, which have a high rate of spontaneous recovery, masking any demonstrable benefit for the severe palsy subgroup. Supporting this are the positive findings in favour of combined therapy in non-double-blinded studies.

The key rationale for giving patients with Bell’s palsy antiviral treatment with the antiherpes drug acyclovir is the possible role of HSV-1, based on the current circumstantial evidence. Another reason given for using antiviral therapy in the setting of clinical equipoise is that a portion of those given a provisional diagnosis of Bell’s palsy will have zoster sine herpete, that is, symptomatic VZV reactivation without the typical vesicular eruption pathognomonic of a typical VZV infection (Ramsay-Hunt syndrome).

VZV is an important differential diagnosis in all acute lower motor neuron facial palsies. Ramsay-Hunt syndrome has a worse prognosis than Bell’s palsy and, on average, presents as a more severe palsy. It is more responsive to combined antiviral and steroid therapy, and the rate of complications from antiviral therapy is low. Since VZV is known to cause facial palsy, the use of antivirals in Ramsay-Hunt syndrome has a clear evidence base, and is justified and rational. A typical regime to adequately cover VZV would be 3000 mg/day (1000 mg valacyclovir three times a day) for 7 days. Valacyclovir has a higher bioavailability than acyclovir.

At the present the use of combined acyclovir and corticosteroids in treating classical Bell’s palsy remains controversial, with conflicting data emerging from different trials and, indeed, from different meta-analyses. Based on current evidence, particularly the extensive Scottish Bell’s palsy study of 551 patients in a double-blind, placebo-controlled, randomised study, it seems reasonable to treat classic Bell’s palsy with oral corticosteroids alone, without acyclovir. However, combined acyclovir and corticosteroids may possibly have a beneficial role in cases of severe Bell’s palsy, and this issue needs to be resolved in a large prospective clinical trial. In cases where the patient is severely immunocompromised, consideration may be given to intravenous regimes of acyclovir to prevent possible central nervous system complications.

Eye care

The institution of an eye protection strategy for each patient with incomplete closure is of paramount importance. Lacrimation occurs at the lateral aspect of the conjunctival membrane and is swept lateral to medial as a film by the action of the orbicularis oculi during blink and effective eye closure. Loss of this mechanism results in epiphora (tearing) due to an ineffective pump mechanism to spread the tear film, and exposure and irritation of the eye itself. Prolonged drying and irritation will progress to keratitis and ulceration, and eventually can threaten sight. At the first consultation, the clinician must enact a strategy to avoid ocular exposure, and refer any cases of concern to an ophthalmologist. Effective eye protection uses barrier protection (eg, wrapped sunglasses), lubrication (artificial tears during the day, ointment at night) and taped closure at night. Particular environments that may present a challenge for the patient include showering and swimming, and dusty and windy environments; these situations are best avoided.

The early use of an eyelid weight should be considered in the elderly, those with diabetes, with pre-existing eye disease, complete facial palsy with no response on neurophysiology and the presence of ongoing irritation despite the use of the eye-protective therapies described above.

Oral care

Loss of the sphincter function of the orbicularis oris confers the social inconvenience of oral incontinence and predisposes the lip and inner cheek to abrasion during mastication and subsequent ulceration. Strategic eating may lessen the impact of these
in the setting of flaccid facial paralysis. Using a straw for liquids and tending towards soft foods are often helpful. The inability to lower and evert the lower lip precludes eating certain foods. Temporary dental ‘spacers’ adhered to the lateral aspect of the molar teeth may be used to prevent chewing of the buccal mucosa.

Physiotherapy
From an evidence-based perspective, this diverse modality of treatment, which broadly encompasses heat therapy, electrostimulation, massage, mime therapy and biofeedback, is hard to analyse as a whole. There is a plethora of treatment regimes, and their timing and variability in implementation make their broader assessment of utility complex. Although not recommended for all sufferers of Bell’s palsy, there are subgroups of patients for whom there is evidence supporting the use of physiotherapy. These include patients who have incomplete recovery and who have developed hypertonia, hyperkinesis or synkinesis, and in these groups neuromuscular retraining is trialled before consideration of chemodenervation.

Prognosis
Natural history studies have demonstrated that ~85% of patients experience some recovery in the first 3 weeks. Predictors of incomplete recovery include severe facial palsy, the length of time prior to onset of recovery, and persistent pain. Patients with complete facial palsy (House-Brackman grades 5–6) who have not experienced some recovery in the first 3–4 months after onset are more likely to have incomplete recovery of facial function, with or without spasm and synkinesis. Prolonged pain is also a predictor of worse outcome.

The natural history of Bell’s palsy has been elucidated through a number of large studies. Based on conclusions drawn from these large studies, clinicians can expect that without intervention, approximately 70% of patients will experience full recovery. Of those who do not recover fully, half will have mild sequelae, and the remainder moderate-to-severe sequelae. In the setting of acute steroid use the rate of full recovery is over 90%.

Managing incomplete recovery
Chronic facial palsy is a disabling condition that has a dramatic impact on social function, emotional expression and quality of life. Aesthetic, functional (nasal patency, eye closure, speech and swallowing) and psychological considerations need to be addressed by the treating team. Over the last three decades, the treatment of incomplete recovery of facial palsy has evolved from static techniques aimed at suspension of the oral commissure and eye closure, into a multimodal, zonal-based approach that utilises the complementary aspects of physiotherapy, chemodenervation and selective surgical procedures to maximise the cosmetic and functional outcome needs of each patient. Increasingly, multidisciplinary collaboration between interested clinicians from a wide variety of subspecialties has proven effective.

In outlining treatment modalities and their appropriateness, one must consider incomplete recovery of facial function as a heterogeneous entity that encompasses different degrees of flaccidity, hypertonicity and synkinesis. Each of these issues can range in severity from absent to severe. Generally, functional issues such as brow ptosis, nasal valvular collapse and eye closure, are addressed through directed structural interventions including nasal valvular suspension, brow ptosis correction, platinum weight insertion into the upper eyelid, lower lid suspension or tarsorrhaphy to improve eye closure.

Synkinesis
Synkinesis refers to abnormal facial muscular contraction during voluntary facial movements and has been traditionally attributed to aberrant re-innervation of facial musculature following nerve injury. It can be seen as involuntary eye-closure during midface movement such as eating or smiling (oro-ocular synkinesis); as lip excursion during eye closure (oculo-oral synkinesis); or as chin dimpling or muscular neck cords during midface movement due to involuntary mentalis or platysma activation. The
treatment of synkinesis centres around physiotherapy, with a particular focus on biofeedback exercises to retrain facial symmetry, and selective chemodenervation using directed botulinum toxin to problem areas (figure 2). Most patients express a high degree of satisfaction with this approach, and objective improvement is seen in the majority.38

Treatment of synkinesis with botulinum toxin is tailored according to the individual patient. Injection points focused on orbicularis occuli and platysma relieve spasms and synkinesis while selective use of contralateral (unaffected side) brow and depressor angularis oris injection points enhances facial symmetry and cosmesis. Injection of the weak/synkinetic zygomaticus muscles is best avoided due to the debilitating effect of losing smile function from an already weakened face. Some patients have a reduced benefit with repeated injections while for others the requirement for recurrent injections three to four times per year is unsatisfactory. In selected cases, a more permanent solution can be achieved through surgical interventions such as selective myectomy. Recently, selective neuroectomy has also proven to be very effective.19

Facial reanimation

In the setting of incomplete recovery with ineffective smile, nerve-to-nerve transfer, and regional and free tissue transfer techniques offer the opportunity to restore midface movement. A regional tendon transfer technique that relocates the temporalis muscle tendon to the oral commissure has been popularised by the French surgeon, Labbé.30 Alternatively, innervated free muscle transfer techniques that utilise the gracilis muscle can be inserted into the face with nerve coaptation from the contralateral facial nerve and/or from the ipsilateral mandibular branch of the trigeminal nerve. Nerve-to-nerve transfers are also becoming increasingly popular for those with autopreservation of facial musculature but a frozen oral commissure. Donor nerves include the masseteric branch of the trigeminal nerve, and cross face nerve grafting. These complex reconstructive procedures offer an opportunity for smile reanimation, and the subsequent benefits in social function and quality of life.

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES