Hypopituitarism following traumatic brain injury: prevalence is affected by the use of different dynamic tests and different normal values

Nieke E Kokshoorn, Moniek J E Wassenaar, Nienke R Biermasz, Ferdinand Roelfsema, Johannes W A Smit, Johannes A Romijn and Alberto M Pereira
Department of Endocrinology and Metabolic Diseases C4-R, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands
(Correspondence should be addressed to N E Kokshoorn; Email: n.e.kokshoorn@lumc.nl)

Abstract

Objective: Traumatic brain injury (TBI) has emerged as an important cause of hypopituitarism. However, considerable variations in the prevalence of hypopituitarism are reported. These can partly be explained by severity of trauma and timing of hormonal evaluation, but may also be dependent on endocrine tests and criteria used for diagnosis of hypopituitarism.

Methods: Systematic review of studies reporting prevalence of hypopituitarism in adults ≥ 1 year after TBI focusing on used (dynamic) tests and biochemical criteria.

Results: We included data from 14 studies with a total of 931 patients. There was considerable variation in definition of hypopituitarism. Overall, reported prevalences of severe GH deficiency varied between 2 and 39%. Prevalences were 8–20% using the GHRH–arginine test (cutoff <9 µg/l), 11–39% using the glucagon test (cutoff 1–5 µg/l), 2% using the GHRH test (no cutoff), and 15–18% using the insulin tolerance test (ITT; cutoff <3 µg/l).

Overall, the reported prevalence of secondary adrenal insufficiency had a broad range from 0 to 60%. This prevalence was 0–60% with basal cortisol (cutoff <220 or <440 nmol/l), 7–19% using the ACTH test, and 5% with the ITT as first test (cutoff <500 or <350 nmol/l). Secondary hypothyroidism was present in 0–19% (free thyroxine) or 5–15% (thyroid-releasing hormone stimulation). Secondary hypogonadism was present in 0–29%.

Conclusion: The reported variations in the prevalence rates of hypopituitarism after TBI are in part caused by differences in definitions, endocrine assessments of hypopituitarism, and confounding factors. These methodological issues prohibit simple generalizations of results of original studies on TBI-associated hypopituitarism in the perspective of meta-analyses or reviews.

European Journal of Endocrinology 162 11–18

Introduction

In recent years, an increasing number of studies have reported the presence of pituitary insufficiency in patients who experienced traumatic brain injury (TBI) (1–14). The prevalence of pituitary insufficiency after TBI appeared to be unexpectedly high (15, 16). Remarkably, the prevalence rates varied considerably among the different studies, ranging from 15 to even 90% of the patients.

Several factors influence the prevalence of hypopituitarism after TBI. First, the time interval between TBI and the assessment of pituitary function, since hormone alterations mimicking pituitary insufficiency are prevalent in the early post-traumatic period. Secondly, the type and severity of the brain injury affect the prevalence of hypopituitarism, because persistent pituitary insufficiency is only frequent after severe TBI (7, 15). Thirdly, endocrine tests, assays, and criteria for the diagnosis of hypopituitarism differ between the studies. Although many reviews have addressed TBI-related hypopituitarism, a detailed comparison of these methodological issues between the different studies has not been performed for each pituitary axis.

We hypothesized that these methodological differences may have contributed, at least in part, to the discrepancies in prevalence rates of hypopituitarism after TBI, reported by the different studies. Therefore, the aim of this study was to critically compare the pituitary function tests, and definitions of hypopituitarism between studies that assessed the long-term outcome of TBI on pituitary function.

European Journal of Endocrinology (2010) 162 11–18

© 2010 European Society of Endocrinology

DOI: 10.1530/EJE-09-0601

Online version via www.eje-online.org
Subjects and methods

Search strategy

We performed a search in PubMed, EMBASE, Web of Science, and the Cochrane database for all published studies on the association between TBI and hypopituitarism. The following search strategy was used: (TBI OR traumatic brain injuries) AND (traumatic OR trauma) AND (hypopituitarism OR hypopituitar* OR hypothalamus–hypophysis system OR 'hypothalamo-pituitary dysfunction' OR 'pituitary dysfunction' OR hypothalamo-hypophyseal system OR pituitary gland OR hypophysis).

In addition, the references of relevant articles were checked for additional articles. The search was performed on 23 March 2009. Only original articles were included. We used the following exclusion criteria: pediatric or adolescent population; publications concerning pituitary testing 12 months after injury (a median of 12 months was accepted); articles that evaluated pituitary insufficiency after subarachnoidal bleeding.

Data review

The following data were extracted from each study: i) age and gender, ii) the endocrine tests used for assessment of each pituitary axis, iii) definitions used for pituitary insufficiency for each pituitary axis, iv) hormone assays, v) reference values provided in the manuscript, and vi) use of control populations. Tables were constructed per pituitary axis. These tables are added as supplemental data files.

Results

We identified 278 articles, of which 218 were excluded on the basis of title and abstract. Of the remaining 60 articles, 46 were reviews. Finally, 14 original studies were included with a total of 931 patients. Details of these studies are summarized in Table 1. The number of patients evaluated by the different studies varied between 22 and 105.

The GH–IGF1 axis

The prevalence of GH deficiency (GHD) ranged between 2 and 66% (severe GHD 39%; Figs 1 and 2 and Supplementary Table 1). The presence of GHD was associated with higher body mass index (BMI) values in some of the studies (see Fig. 1). In addition to basal...
serum GH and IGF1 values, all studies used a dynamic test to assess GH-secretory reserve. However, different dynamic tests were used.

Three studies (196/931 = 21% of all patients) used the combined GHRH–arginine test as the first screening. The criterion for severe GHD was a peak GH level < 9.0 μg/l in all three, which was not adjusted for BMI. Prevalence rates of severe GHD varied between 8 and 20% (weighted mean 12%) (2, 3, 5). Schneider et al. (11) also used the GHRH–arginine test, but only in a subset of the patients (those with abnormal serum cortisol levels, n = 32); the prevalence of GHD in this study was 10%.

Two studies (112/931 = 12% of all patients) used an insulin tolerance test (ITT) as the primary screening test (6, 7). The criterion for severe GHD was a peak GH response < 3 μg/l in both, and the prevalence of GHD was comparable (18 and 15% respectively; weighted mean 16%).

Of the eight remaining studies, three used a stimulation test with glucagon (n = 209) (1, 4, 9) with prevalence rates for severe GHD between 11 and 39% (weighted mean 20%). The cut-off values differed considerably and varied between 1 and 5 μg/l between these studies. Just one study used a stimulation test with GHRH only (number of patients not recorded) reporting a GHD prevalence of 2% (13). Two studies (n = 119) used the combined GHRH + GHRP-6 test with a prevalence of 15 and 33% respectively (weighted mean 21%) (10, 12). The cut-off values were similar (GH < 10 μg/l) within these studies, and were derived from another report (17).

Finally, two studies used a combination of these tests (8, 14). For instance, Agha et al. (1) used a glucagon stimulation test for the initial screening in 102 subjects, and in case of incomplete GH response, they used an ITT (n = 14) or combined GHRH + arginine test (n = 4) to confirm GHD.

The pituitary–adrenal axis

The prevalence of secondary adrenal insufficiency deficiency ranged from 0 to 60% between the studies (Fig. 3 and Supplementary Table 2).

Four studies (251/931 = 27% of all patients) only measured basal morning fasting serum cortisol and/or ACTH levels (2–4, 10), resulting in prevalence rates between 0 and 60% (weighted mean 15%). The criteria for pituitary–adrenal insufficiency differed between three studies (cortisol < 220–440 nmol/l), and were not reported in the fourth study (10). The study reporting the highest prevalence of 60% used a cut-off value of 440 nmol/l (4).

Four studies (145/931 = 12% of all patients) used an ACTH stimulation test (synacthen; either with 1 or 250 μg). However, only one study performed this test in all patients and the prevalence of ACTH deficiency was 7% (9). In the other three studies, only a subset of the patients (those with subnormal basal cortisol levels) underwent stimulation with ACTH. The prevalence in these studies varied between 7 and 19% (weighted mean 10%) (11, 12, 14).

One study (55/931 = 6% of all patients) used non-stimulated cortisol values between 1600 and 2000 h (reference values 63–339 nmol/l), which was followed by a corticotrope-releasing hormone (CRH) test only in those with values below this reference range, or in those who responded confirmatory to a specific questionnaire (13).

In the remaining five studies (403/931 = 43% of all patients), the ITT was used in 169 patients as a primary test (n = 112) resulting in a prevalence of 5% in both studies (6, 7), or as a confirmation test in a subset of the patients. Two studies measured basal serum cortisol levels and used ITT as a confirmation test (prevalence of 3 and 11% respectively) (5, 8). One study assessed primarily with a glucagon stimulation test (n = 102), and used the ITT and ACTH tests to confirm...
ACTH deficiency (prevalence 13%) (1). The criteria for a normal cortisol response to hypoglycemia were a peak cortisol level >550 nmol/l in one (8), and >500 nmol/l in three other studies (1, 5, 7). The fifth study used a control group of 18 healthy subjects to define normal cortisol responses to ITT (cortisol response >95% confidence limit according to the obtained area under the curve) (6). The CRH test was used in only one study and did not report the number of patients (13).

**The hypothalamus–pituitary–thyroid axis**

The prevalence of hypothalamus–pituitary–thyroid axis deficiency ranged from 0 to 19% between the studies (Supplementary Table 3).

The criteria for TSH deficiency were different (see Supplementary Table 3). Nine studies used basal free thyroxine (FT4) and TSH levels only. Within these studies, the cut-off value for decreased FT4 varied between 8 and 12 pmol/l (2, 5, 7, 11, 14, 18, 19). In two studies, reference values were not reported (4, 10), one of which (Bushnik et al.) reported the highest prevalence of secondary hypothyroidism.

The thyroid-releasing hormone (TRH) stimulation test was used in five studies, using i.v. doses of 200 (13) and 500 µg (5, 6, 8, 9). The criterion for a normal response differed considerably: a TSH peak response >7 mIU/l, a TSH peak between 5 and 30 mIU/l, or were not reported (12, 13).

The prevalence rates between the studies that only measured basal FT4 levels varied between 0 and 19% (weighted mean 5%) (1–5, 7, 10, 11, 14), and between 5 and 15% (weighted mean 8%) in those that also used TRH (6, 8, 9, 12, 13).

![Figure 2](image1.png) Absolute and weighted mean prevalence rates of severe GH deficiency (GHD) according to the stimulation tests used per study. The number of patients tested is depicted in each bar. Panel A, the combined GHRH–arginine test; definition severe GHD: peak GH <9 mIU/l for all four studies. Panel B, the insulin tolerance test (ITT); *definition severe GHD: GH <95% CL according to AUC; **definition severe GHD: peak GH <3 µg/l. Panel C, the combined GHRH–GHRP-6 test; definition severe GHD: peak GH <10 µg/l for both studies. Panel D, the glucagon stimulation test; definition severe GHD: *peak GH <3 µg/l; **peak GH <5 µg/l. Panel E, combined stimulation tests as initial screening followed by confirmation test; *GHRH + GHRP-6 test as initial test; ITT and glucagon stimulation test as confirmation tests; **GHRH + GHRP-6 test as initial test; glucagon stimulation test as confirmation test.

![Figure 3](image2.png) Absolute and weighted mean prevalence rates of corticotropin (HPA axis) deficiency according to the stimulation test used per study. The number of patients tested is depicted in each bar. Panel A, basal cortisol concentrations only using different cut-off levels; *cut-off level: NR; **cut-off level: cortisol <220 nmol/l; ***cut-off level: cortisol <440 nmol/l. Panel B, the ACTH stimulation test *using 250 µg ACTH and peak cortisol <500 nmol/l; **using 1 µg ACTH and peak cortisol <550 nmol/l. Panel C, the insulin tolerance test (ITT); *peak cortisol <95% CL according to AUC; **peak cortisol <500 nmol/l. Panel D, other stimulation tests: the glucagon stimulation test; **CRH test. NR, not reported.
The hypothalamus–pituitary–gonadal axis

The hypothalamus–pituitary–gonadal axis deficiency ranged from 0 to 29% (weighted mean 13%) between the studies (Supplementary Table 4). Basal LH and FSH were measured in all but one study (4). Basal estradiol (E2; in women) was measured in 9 studies, and the menstrual history was recorded in 10 out of 14 studies. Testosterone (in men) was measured in all studies. In four studies, a GnRH stimulation test was performed in a subset of the patients (6, 8, 9, 13). The criterion for a normal test response differed between the studies (see Supplementary Table 4). The definition of secondary hypogonadism was mainly based on basal testosterone (in men) and E2 concentrations (in women) below the reference ranges, in the presence of decreased or normal LH and FSH levels. A subset of the studies also incorporated the GnRH test result (see above) and menstrual cycle abnormalities in premenopausal females.

Prolactin

The prevalence of abnormal serum PRL concentrations ranged from 0 to 16% (Supplementary Table 5). Abnormal PRL secretion was defined as hyperprolactinemia (8/14 studies) (1, 2, 5, 7, 9, 11, 12, 14), hypoprolactinemia (1 study) (6), or both (3). In accordance, prevalence rates were between 3 and 12% using the definition of hyperprolactinemia, 0% using the definition of hypoprolactinemia, and 16% using the combination of both. Out of the 14 studies, 10 measured basal serum PRL concentrations only (1–3, 5, 7, 9–12, 14). Three studies also used a TRH test (doses 100 and 500 µg respectively) (6, 8, 13). Prevalence rates were not reported in two of these (8, 13) and were 0% in the third (6).

Discussion

This review demonstrates that the endocrine evaluations and definitions of hypopituitarism differ considerably among the studies that have assessed TBI-related hypopituitarism. From the existing literature, the notion emerges that most of the tests that are currently used to establish the diagnosis of hypopituitarism in general, and GHD in specific, are not validated sufficiently regarding cut-off values, reproducibility, and dependence on confounding factors in TBI patients. In general, there are hardly any data on reproducibility of tests or dependence on confounding factors in TBI patients. One factor that comes forward in the current review is the potential effect of increased BMI, which in general is associated with decreased GH responses to GH stimulation tests. Therefore, increased BMI may result in an inadvertently higher incidence rate of GHD, if the cut-off values for normal GH responses to GH stimulation tests are not adapted according to BMI. All these methodological issues limit the applicability of the individual studies, i.e. the decision whether the study results are valid for patients to whom the results are generalizable but who are subjected to a different endocrine diagnostic assessment than the original study population. Moreover, these methodological limitations prohibit simple generalizations of the results from the perspective of a meta-analysis or a review.

The question arises whether post-traumatic hypopituitarism, especially GHD, has been over-diagnosed on the basis of the older cross-sectional studies. Consensus guidelines for the evaluation of adult GHD state that different dynamic tests can be used to diagnose GHD, including the ITT, the glucagon stimulation test, the combined GHRH–arginine test, and the combined GHRH–GHRP-6 test (20). The present assessment, however, documents a higher prevalence of GHD for the glucagon stimulation test and the combined GHRH–GHRP-6 test, compared to the results of the combined GHRH–arginine test and the ITT. With the exception of the ITT (which was used in only 12% of the patients (6, 7)), the outcome of each test varied greatly (see Fig. 2) using different (glucagon stimulation test) or similar cut-off levels (other test). In addition, the studies that used two dynamic tests to assess GH reserve revealed a lower prevalence of GHD than the studies with only one test. Moreover, the results of GH stimulation tests are confounded by BMI, with higher BMI being associated with decreased GH responses. Although BMI-adjusted reference values have been reported (21), none of the studies on TBI-associated GHD reports adjusted their cut-off values for BMI. Moreover, the data indicate that BMI tends to be higher in the TBI patients with GHD (see Fig. 1). Finally, an important aspect is that most patients had only GHD or one additional pituitary hormone deficiency. Therefore, one test may not be sufficiently reliable and the use of two tests would increase the confidence in the diagnosis, although only if the two tests yield concordant results. However, the application of two tests may introduce an even greater uncertainty in case of discordant results. This discrepancy has been documented for instance in GHD in irradiated patients, in whom the attenuation in GH responses to the ITT was greater compared with the combined GHRH–arginine test (22). These factors impose major problems for an accurate assessment of GHD in these patients. Therefore, these methodological issues have contributed to the suspicion that GHD is probably over-diagnosed in the older cross-sectional studies.

We observed similar variations in test results of the pituitary–adrenal axis (see Fig. 3). The use of different tests with different cut-off values resulted in prevalence rates that varied between 5 and 19%. ACTH deficiency can be diagnosed by measuring basal early morning cortisol levels: values below 100 nmol/l are indicative of ACTH deficiency, whereas cortisol values above 500 nmol/l essentially exclude ACTH deficiency.
The ACTH stimulation test is reliable in diagnosing clinically significant adrenal insufficiency in patients who are at risk (23–25). ACTH stimulation tests, however, are not fully reliable in excluding the presence of mild secondary adrenal insufficiency (26). The ITT still remains the golden standard, and has the advantage that ACTH/cortisol and GH secretory reserve can be assessed simultaneously. If an ITT is contraindicated, a CRH test can be alternatively used (27). The effect of the initial choice for a specific stimulation test on the variation in outcome of adrenal insufficiency and GHD based on the available data after TBI is illustrated in Fig. 4.

The diagnosis of secondary hypothyroidism is usually made based on FT₄ values. However, basal FT₄ levels show a relatively small intra-individual variability, although inter-individual variability is large (28). As a consequence, a diagnosis of possible secondary hypothyroidism may not be straightforward, since FT₄ levels within the normal reference range can reflect hypothyroidism in one patient but euthyroidism in another patient. Basal TSH levels are also of limited help for the diagnosis of secondary hypothyroidism, since normal or even increased levels of TSH can be found (29). In addition, a TRH test is of limited value because patients with central hypothyroidism may show different patterns of TSH responses to TRH, with absent or exaggerated responses, which considerably overlap with those found in healthy volunteers. Moreover, the magnitude of the TSH peak is proportional to the injected TRH dose, is higher in women, and tends to decline with age (30). In accordance, the prevalence rates were not affected by the use of TRH stimulation. In analogy, the interpretation of the GnRH test is complex, and individual responses vary greatly in both adults and children (31). In men, it is sufficient to measure non-stimulated LH, FSH, and testosterone concentrations. In premenopausal women, the evaluation of the menstrual cycle is a prerequisite, whereas in postmenopausal women, the absence of increased LH and FSH levels almost invariably indicates hypogonadotropic hypogonadism.

Analytical factors will most likely also have affected the different outcomes of the studies. For instance, the GH and cortisol assays varied between studies, and it is known that the between-laboratory performance of the GH assay is not very good. Moreover, most were not validated sufficiently regarding normal cut-off values, reproducibility, and dependence on confounding factors even in a ‘normal’ population. None of these tests have been validated in TBI patients at all.

The time point of evaluation may also influence outcome; therefore, we focused only on studies in the chronic phase after TBI, i.e. 1 year after the trauma. Studies that analyzed patients with a median duration of 12 months after TBI, however, were also included. Thus, part of these patients was assessed within 12 months after TBI. In general, the transient effects of TBI mimicking pituitary insufficiency are almost exclusively reported only within the first 6 months after TBI (15). Therefore, it is unlikely that the pituitary results of the studies with a median duration of followup of 12 months of TBI are caused by the transient effects of TBI. This is supported by similar results of additional analyses of the remaining studies, which included only patients with a followup of more than 12 months after TBI. Lastly, the underlying mechanisms of TBI-related hypopituitarism have not been resolved. It is unclear to what extent hypothalamic versus pituitary damage is present in TBI patients with hypopituitarism and what impact these processes may have on endocrine tests.
Recently, many clinical reviews have summarized the studies on pituitary insufficiency after TBI (15, 16). These studies concluded that hypopituitarism is a common complication of TBI and might contribute to morbidity and poor recovery after brain injury (16). However, these reviews did not take into account the variability in diagnostic strategies and definitions of pituitary insufficiency. These discrepancies, in addition to differences in inclusion and exclusion criteria, limit the possibility to compare the results of studies on TBI. We agree with Klose & Feldt-Rasmussen that future studies should be designed to ensure a high diagnostic robustness for proper identification of reliable predictors, as the results may be highly dependent on diagnostic pitfalls (15).

In conclusion, the reported prevalence rates of pituitary insufficiency after TBI vary considerably, which is associated with major differences in endocrine and analytical methods of assessment and definitions used for hypopituitarism. This does not only apply to the case of TBI-related hypopituitarism, but most likely also to hypopituitarism caused by pituitary diseases. The same caution with respect to the evaluation of pituitary function should be considered in pituitary diseases, because the diagnosis of definitive hypopituitarism remains a challenge in clinical endocrinology. In pituitary pathology, definitive data on robust accuracy of basal or dynamic hormonal tests are incomplete.

Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/EJE-09-0601.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This research did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sector.

References


28 Andersen S, Bruun NH, Pedersen KM & Laurberg P. Biologic variation is important for interpretation of thyroid function tests. *Thyroid* 2003 **13** 1069–1078.


31 Westwood ME, Butler GE, McLellan AC & Barth JH. The combined pituitary function test in children: an evaluation of the clinical usefulness of TRH and LHRH stimulation tests through a retrospective analysis of one hundred and twenty six cases. *Clinical Endocrinology* 2000 **52** 727–733.