

Investigating the real reason for the misery of 'fit for work' assessments

by Kaliya Franklin

Introduction

The Work Capability Assessment (WCA) used to determine eligibility for Employment and Support Allowance (ESA) has been dogged with problems since its inception. Department for Work and Pensions (DWP) ministers and Atos have always denied the existence of 'targets' for the WCA. Now, for the first time, this article presents evidence that the WCA operates to a norm-referenced system. This is a de facto target system, since when 'statistical norms' are applied to a process such as the WCA they deliver the same outcome as targets. As Lord Boswell predicted in 2007, this use of 'statistical norms' means "the test will, in effect, be geared to deliver that [desired] result." Being able to consistently deny the existence of any targets has been crucial to the continuing use of the WCA.

What's wrong with 'statistical norms'?

The media regularly feature stories of people even the Daily Mail would consider deserving of ESA who have been refused the benefit – people with sickness or disabilities as diverse as Huntington's Disease, uncontrolled epilepsy, kidney failure or brittle bone disease. Why are so many of these people being failed by the WCA and who is responsible for that failure? Ministers, MPs, the Department for Work and Pensions and campaigners point the finger at Atos Healthcare, the company contracted to carry out these tests on behalf of the DWP. The DWP say Atos are at fault and that they will be improving the system by breaking the monopoly provision and allowing other companies to bid for the contract once it is due for renewal next year. Relationships between politicians, the DWP and Atos are at an all-time low and now Labour have also announced that they too intend to be rid of Atos. The DWP ministers blame the civil servants and the civil servants seethe. It is a classic case of political, "it wasn't me Miss, it was them".

In 2007 the dangers of introducing a norms-based system to ESA assessments were highlighted during a debate in the House of Commons by Timothy Boswell, the then Conservative MP for Daventry and now Lord Boswell of Aynho. Lord Boswell explained that Incapacity Benefit used a criterion referenced system and warned:

"I can imagine circumstances...in which a future Minister, under financial or even political pressure, might wish to say, **'We will introduce a norm. We are not going to have, by definition, more than 1.5 million people on employment and support allowance,' and the tests will, in effect, be geared to deliver that result.**"

Timothy Boswell MP, 9th January 2007¹

¹ <http://www.publications.parliament.uk/pa/cm200607/cmhansrd/cm070109/debtext/70109-0006.htm>

The essential feature of norm-referencing is that it awards a 'score' on the basis of the ranking of that case within a defined cohort or group. In contrast, criterion-referencing awards a 'score' on the basis of comparing achievement with clear, objective measures. It is easiest to explain how this works in the context of marking student exams. If a criterion-referencing system is used to mark scores, students are given an objective grade based on their performance against the pre-defined marking scheme. This means that in a criterion system if the mark for an 'A' grade is 70%, every student who scores 70% or more will receive an 'A' grade.

However, if a norm-referenced system is used then students' criterion-based scores are ranked in order within their cohort, typically their class or year. This means that in a norm-referenced system there is effectively a 'target' for the proportion of 'A' grades to be awarded so only those students who **both** score above 70% **and** fall within the allowed number of 'A' grades will receive an 'A' grade.²

As mentioned above, the test for Incapacity Benefit, the Personal Capability Assessment, was a criterion-based system. People were awarded points based upon how they scored against certain criteria – for example, those who could walk less than 50 metres were awarded more points than those who could walk less than 200 metres. If the person scored the number of points the criteria demanded for benefit receipt then they would be entitled to the benefit.

However, the Work Capability Assessment used to consider eligibility for Employment and Support Allowance, the replacement for Incapacity Benefit, is a norm-referenced system. People must **both** score the number of points required for benefit receipt **and** fall within the proportion of people the norms system will allow to receive the benefit. In practice this means there is a finite number of claims the assessment system will allow to be awarded the benefit, regardless of the number of people who objectively meet the criteria for benefit eligibility.

Atos, norms and the blame game

So, is removing Atos actually a solution to the problems with the WCA? Would replacing Atos with Capita, Serco or G4S really improve outcomes for those sick and disabled people who require the financial support provided by ESA? Would it reduce the cost to the taxpayer of successful appeals or the human suffering caused when seriously sick or disabled people are denied any financial support? The answer, sadly, is no. As long as the WCA remains a norms based system of assessment, removing Atos is the political equivalent of fiddling while Rome burns – it seems to satisfy the demands of campaigners as it is seen as 'doing something' while, quietly and behind the scenes, the fire continues to rage.

The WCA was conceived from a report commissioned under the previous Labour government, prepared by the then banker, now Minister for Welfare Reform, Lord David Freud.³ Despite admitting he knew nothing about welfare or disability, within three weeks Freud concluded that some two thirds of those receiving Incapacity Benefit (the predecessor to Employment and Support Allowance) were indeed capable of work. And so it began: a system designed not to assess whether someone was really too sick or disabled to work- i.e. not the 'real world' test Chris Grayling was so

² <http://www.publications.parliament.uk/pa/cm200607/cmhansrd/cm070109/debtext/70109-0006.htm>

³ [http://base-uk.org/sites/base-uk.org/files/\[user-raw\]/11-07/welfarereview.pdf](http://base-uk.org/sites/base-uk.org/files/[user-raw]/11-07/welfarereview.pdf)

“implacably and unreservedly opposed to” when he was Minister for Employment⁴ - but a system designed to allocate as few points as possible to an individual by ignoring real world barriers to work and assessing a person’s ability to ‘function’ against the most basic and irrelevant descriptors such as lifting an empty cardboard box. Over the years there have been improvements made to the WCA – some of the descriptors have been removed, a personal statement intended to explain to the claimant why they have or have not been allocated the benefit was added after recommendations from the independent reviews into the WCA run by Professor Harrington.⁵ Audio recording of the assessments is also theoretically permitted, but in practice prevented by lack of equipment.

However, it is clear that this is more fiddling, or changing the mood music, as the rate of successful appeals against ESA decisions continues to rise:

Appeal outcomes:

For the year 2012-2013 the appeal statistics from the Ministry of Justice show the following:

327,961 appeals were received (table 1.4)

268,137 appeals were “disposed” (table 2.6)

Of these disposals 42,901 (16%) were "cleared without a hearing" (including cases that were struck out, superseded or withdrawn before hearing, table 2.6), so

225236 appeals were heard (table 2.7), of which approximately 43% were decided in favour of the claimant⁶.

Cost of appeals:

The cost of ESA appeals to HM Courts and Tribunal Service for that same period (April 2012-March 2013) was £66 million⁷. In addition the cost to the DWP of all activity relating to those same appeals was £17.1 million⁸. Therefore the **total** cost to the public purse of ESA appeals between April 2012 and March 2013 was £83.1 million.

Why is there such a great disparity between the initial decision by DWP and the decision made by the tribunal service managed by the Ministry of Justice? Also, why is the government delaying the

⁴ Employment Minister, Chris Grayling, in evidence to the Work and Pensions Select Committee on IB to ESA migration, Wednesday, 8th June 2011, responding to question 340:

<http://www.publications.parliament.uk/pa/cm201012/cmselect/cmworpen/1015/11060802.htm>

⁵ <https://www.gov.uk/government/publications/work-capability-assessment-independent-review-year-1>

<https://www.gov.uk/government/publications/work-capability-assessment-independent-review-year-2>

<https://www.gov.uk/government/publications/work-capability-assessment-independent-review-year-3>

⁶ Tribunal service statistical tables, April-June 2013: <https://www.gov.uk/government/publications/tribunal-statistics-quarterly-april-to-june-2013> (the figures don’t correlate exactly, probably due to rounding)

⁷ Helen Grant MP in answer to a written question from Liam Byrne MP, 17 July 2013:

http://www.publications.parliament.uk/pa/cm201314/cmhansrd/cm130717/text/130717w0002.htm#130717w0002.htm_snew55

⁸ Lord Freud, (House of Lords debate) 18 July 2013, c141W: <http://www.theyworkforyou.com/wrans/?id=2013-07-18a.141.3&s=lord+freud+work+capability+assessment#g141.5>

removal of Atos, who are frequently alleged to be the cause of the problem? The second question is much simpler to answer than the first – to break the contractual terms and sack Atos now would incur huge financial penalties, there is no other provider currently in a position to provide the service (any company which does successfully bid for the contract will have to draw from Atos’s current pool of employees to find enough staff to do the job) and crucially the transfer of those still on the old Incapacity Benefit is not due to be completed until March 2014. These are the longer-term sick or disabled people, the group from which Freud was convinced he could find two thirds fit for work. Two thirds of the people in this group are also entitled to Disability Living Allowance - which uses both more sensible and more realistic criteria to assess a person's level of disability. This is a group of people with more significant level of sickness or disability than those claiming ESA for the first time - these new claimants are more likely to have the kind of illness or injury which lasts months rather than years⁹.

Back to the first question – why is there such a disparity between the decisions made by DWP decision makers, based on assessments provided by Atos performing to a DWP contract, and the decisions made by the tribunal service, consisting of a doctor, a lawyer and a disabled lay member not performing to a DWP imposed contract? The answer is both depressingly simple and hideously complicated.

‘Norms’

An example from education helps explain how ‘norm-referencing’ results in a ‘quota’ or ‘target’. Exams can be either ‘norm-referenced’, ‘criterion-referenced’ or a mix of both. A criterion-referenced exam system means students receive an objective grade based on their performance against a pre-defined marking scheme. A norm-referenced exam system sets ‘quotas’ which limit the overall proportion of students able to achieve each grade.

Originally ‘A’ levels were simply pass or fail, but in 1963 guidance was issued which limited the overall proportion of students allowed to receive each grade.¹⁰ This was a norm-referenced system which ensured that only a small proportion of students could achieve ‘A’ grades. For some exams the difference between a ‘B’ grade and a ‘D’ grade was less than 10 marks. This was widely perceived as unfair to students and during the 1980’s ‘A’ level marking systems were changed to remove quotas.

The use of ‘statistical norms’, or expectations, to govern decisions on which sick and disabled claimants are awarded benefit, goes a long way towards explaining why attempts by politicians, charities and campaigners to achieve positive change to the WCA have had such little effect. The next section examines in detail a key piece of evidence, which shows how the prescribed ‘statistical norms’ have been incorporated into the ‘management information tool’ used by Atos:

⁹ Work and Pensions Longitudinal Study via Nomis

¹⁰ <http://www.publications.parliament.uk/pa/cm200203/cmselect/cmmeduski/153/15304.htm>

MSC	Total Number of Assessments	Total Number of Referrals	Below Normal Range		Above Normal Range		20%		-20%		-20%		-20%		-20%		-20%		-20%		-20%		-20%	
			% In Support Group	% Curtailed	% Support Group Or Curtailed	% NFD	% 3m Prognosis	% 6m Prognosis	% 12m Prognosis	% 18m Prognosis	% 2y Prognosis	% Longer Term Prognosis	Average Physical Score	Average Mental Score	Average PSS Word Count	Average Time Taken								
Birmingham	5,098	1,929	12.9%	15.6%	28.5%	0.1%	71.3%	4.0%	11.7%	7.1%	3.1%	2.8%	1.8	2.4	191.4	52.6								
Bristol	2,591	825	14.0%	23.8%	37.7%	0.0%	59.9%	11.3%	17.3%	5.1%	3.0%	3.5%	2.2	4.8	164.2	48.8								
Cardiff	2,544	1,036	12.7%	21.8%	34.6%	0.2%	64.0%	6.5%	16.0%	9.3%	2.9%	2.3%	2.4	4.6	200.8	48.9								
Scotland	4,433	1,670	14.1%	19.3%	33.4%	0.1%	66.0%	6.7%	13.7%	6.1%	3.8%	3.7%	1.8	3.2	198.9	46.4								
Leeds	4,604	1,656	16.8%	22.0%	38.8%	0.0%	61.9%	8.1%	16.3%	5.7%	3.3%	2.7%	2.0	3.4	192.2	45.9								
Manchester	4,172	1,505	13.4%	24.9%	38.1%	0.1%	62.2%	9.4%	17.4%	6.4%	2.8%	2.7%	2.2	4.2	196.2	50.7								
Newcastle	3,035	1,072	13.6%	14.2%	27.8%	0.1%	73.0%	4.2%	10.2%	5.4%	3.8%	3.3%	1.4	2.3	152.3	49.2								
Nottingham	5,069	2,222	13.9%	22.7%	36.7%	0.0%	63.3%	9.9%	17.5%	4.4%	2.4%	2.5%	2.5	4.1	232.6	48.1								
Croydon	2,734	894	16.9%	24.3%	41.3%	0.1%	58.1%	11.1%	20.7%	4.8%	3.1%	2.3%	2.2	4.4	219.5	55.6								
Wembley	7,427	2,840	12.2%	19.1%	31.2%	0.1%	67.4%	6.8%	16.7%	5.6%	2.1%	1.3%	2.2	3.4	193.6	49.3								
National	45,162	17,342	14.6%	20.5%	35.0%	0.1%	64.8%	7.5%	15.9%	6.1%	3.0%	2.6%	2.1	3.6	196.5	49.5								

Under the regulations governing ESA, a person must be awarded fifteen points overall to reach the threshold of sickness or disability set by Government to be considered ‘deserving’ enough to receive Employment and Support Allowance for the Work Related Activity Group, and fifteen points on any one descriptor to be eligible for the Support Group. With the ‘statistical norms’ effectively driving down the average number of points scored for physical conditions to just 2.1, and for mental health conditions to 3.6, it is obvious that most claimants will be much less likely to receive the points needed to qualify for the benefit. This is exactly the situation Lord Boswell warned about in 2007, that to impose a norm-referenced system would mean

“the test will, in effect, be geared to deliver that result.”

It is almost unbelievable that the Government persists in its view that a theory devised by a banker, with no knowledge of disability or social security, is an appropriate basis on which to decide whether sick and disabled claimants have an income while they are unable to work.

Are the lives of sick and disabled people in the UK worth so little to those on whom we depend for a safety net?

Detailed analysis of the available evidence

Like many large companies Atos monitor the performance of their assessors and quality of their assessments by using a computer system which interprets the data inputted into the Atos software (Lima) by the practitioner carrying out ESA assessments. The average results for each practitioner are compared with the average results of other practitioners in both the same region and nationally. These criteria for the audit system are set by the DWP and the numbers for the ‘norms’ are given to Atos by the DWP. The norms are based on the percentage of cases the DWP expects to fall into each category. Due to the constraints of the audit system and how it drives behaviour, doctors, nurses or physiotherapists are not truly free to apply their professional judgement as to whether an individual is really fit enough to work. Instead the assessors are constrained by a rigid audit system to ensure all descriptor choices fall within the narrow margin of the norms permitted by the DWP.

In its original contract with Atos Healthcare to deliver the WCA, DWP specified that it expected about 11% of ESA claimants to go into the support group. This is what the contract says:

'It is estimated that approximately 11% of new customers will fall into the support group. Atos Healthcare **MUST** base their solutions and costs the [sic] figures in Appendix 8.'¹¹

This is the statistical result the original WCA process delivered until the initial Harrington review¹², after which the proportion of people being allocated to the Support Group rose to 16% in May 2011. Although these are described as 'statistical norms', to many people this will sound like a target system, especially given the clarity of the language specified in the original contract, that Atos Healthcare

"MUST base their solutions and costs [on the] figures in Appendix 8."¹³

In practice the norms system operates like a de facto target system in two ways. Firstly it guides the outcome of the entire process. Secondly, although in theory the audit is not supposed to be invasive or punitive, in reality it drives human behaviour – assessors fear being placed on '100% audit' by Atos, a practice a bit like having your scariest old teacher standing over you with a red pen and giving you a failing grade for putting a comma in place of a semicolon. The consequences for assessors of continuing 100% audit can be deeply stressful, in some cases affecting the performance-related elements of their pay, and could ultimately result in dismissal. The former Atos assessor Dr Greg Wood said;

"I got the impression that assessors always had one eye on what the auditors would think of their reports. No one wanted to be put on 'the naughty step', as being subjected to 100% report-auditing was known within Atos. It is time-consuming to write the long-winded exam-grade reports expected by the auditors, so you see fewer cases than the daily target of 8, which can lead to managerial scrutiny in itself - double trouble!"

What this audit process does not do is improve the outcome of the assessment, allow for professional judgment to be applied or represent natural justice either for the claimant or the taxpayer.

The management information tool

MSC	Below Normal Range		20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
	Total Number of Assessments	Total Number of Referrals	% In Support Group	% Curtailed	% Support Group Or Curtailed	% NFD	% 3m Prognosis	% 6m Prognosis	% 12m Prognosis	% 18m Prognosis	% 2y Prognosis	% Longer Term Prognosis	Average Physical score	Average Mental Score	Average PSS Word Count	Average Time Taken			
Birmingham	5,098	1,929	12.9%	15.6%	28.5%	0.1%	71.3%	4.8%	11.7%	7.1%	3.1%	2.8%	1.5	2.4	191.4	52.6			
Boole	3,455	1,593	22.3%	21.1%	43.5%	0.1%	58.9%	9.0%	16.1%	8.9%	3.6%	3.4%	2.4	3.5	202.3	50.5			
Bristol	2,591	825	14.0%	23.8%	37.7%	0.0%	59.9%	11.3%	17.3%	5.1%	3.0%	3.5%	2.2	4.8	164.2	48.8			
Cardiff	2,544	1,036	12.7%	21.8%	34.6%	0.2%	64.0%	6.5%	16.0%	9.3%	2.9%	2.3%	2.4	4.6	200.8	48.9			
Scotland	4,433	1,670	14.1%	19.2%	33.4%	0.1%	66.0%	6.7%	13.7%	6.1%	3.8%	3.7%	1.8	3.2	198.9	46.4			
Leeds	4,604	1,656	16.8%	22.0%	38.8%	0.8%	61.9%	8.1%	18.3%	5.7%	3.3%	2.7%	2.0	3.4	192.2	45.9			
Manchester	4,172	1,505	13.4%	24.7%	38.1%	0.1%	62.2%	8.4%	17.4%	6.4%	2.8%	2.7%	2.2	4.2	196.2	50.7			
Newcastle	3,035	1,072	13.6%	14.2%	27.8%	0.1%	73.0%	4.2%	10.2%	5.4%	3.8%	3.3%	1.4	2.3	152.3	49.2			
Nottingham	5,069	2,222	13.9%	22.7%	36.7%	0.0%	63.3%	9.9%	17.5%	4.4%	2.4%	2.5%	2.5	4.1	232.6	48.1			
Croydon	2,734	994	16.9%	24.3%	41.3%	0.1%	58.1%	11.1%	20.7%	4.8%	3.1%	2.3%	2.2	4.4	219.5	55.6			
Wembley	7,427	2,840	12.2%	19.1%	31.2%	0.1%	67.4%	6.8%	16.7%	5.6%	2.1%	1.9%	2.2	3.4	193.6	49.3			
National	45,162	17,342	14.5%	20.5%	35.0%	0.1%	64.8%	7.5%	15.9%	6.1%	3.0%	2.6%	2.1	3.6	196.5	49.5			

¹¹ Contract between DWP and Atos Healthcare, Request for Proposal, section 2.2.2, Appendix 8, shows DWP's modelling of the caseload, including the 11% estimate for the support group: <http://www.docstoc.com/docs/126689560/DWP-to-Atos-Contract-2010> (Note: the link states 2010 but the contract is dated 2007). See <http://lartsocial.org/DLAreform>

¹² <https://www.gov.uk/government/publications/work-capability-assessment-independent-review-year-1>

¹³ See note 12 above

For the first time evidence has emerged of a management information tool, which is a key element of this statistical audit system and makes it clear why none of the changes to the WCA have actually made much improvement.

The colours are a basic ‘traffic light’ system – green indicates that the result is within acceptable range of the norm, blue means it has gone too far below the imposed norm and red means it has gone too far above the allowed norm. The figures above indicate the system that Atos use to actively ‘compare’ the output of each individual practitioner. Deviation from a narrow range of ‘averages’, (deviation from the ‘norm’), is not tolerated and leads to a ‘target audit’ process on that individual practitioner. This leads to practitioners closely monitoring the percentage of cases they have placed into various ‘outcomes’. Once they have placed, say, 15% of claimants into the support group in a week they are very much less likely to put the next seriously ill or severely disabled claimant they see that week into the deserved support group, as to do so would take the practitioner’s weekly figures away from the ‘norm’ and would be very likely to draw the wrath of their manager upon them through the medium of oppressive and intimidating 100% audit.

Another Atos employee (name withheld) stated,

“I feel guilty for doing it but I need my job with Atos to pay my mortgage and feed my children. I have been overtly threatened with ‘re-training’ and dismissal when my support group figures strayed above 20% - after I saw a run of very disabled people in a row one week. It is very depressing to be on 100% audit- it forces you to slavishly follow the rules, rather than being fair to genuinely disabled people.”

Column-by-column analysis of the management information tool

The first column is a list of the regional medical examination centres (MSC). Column two is self-explanatory; total number of face-to-face assessments performed within the month of data displayed – this is the total of both new claims for ESA and those transferring from the older Incapacity Benefit. Column three, total number of re-referrals, refers to the IB transfer group, which is those who are longer term sick or disabled.

MSC	Total Number of Assessments	Total Number of Re-referrals	Below Normal Range		Above Normal Range		20%		20%		20%		20%		20%		20%		20%		20%		20%		20%	
			% In Support Group	% Curtailed	% Support Group Or Curtailed	% NFD	% 3m Prognosis	% 6m Prognosis	% 12m Prognosis	% 18m Prognosis	% 2y Prognosis	% Longer Term Prognosis	Average Physical score	Average Mental Score	Average PSS Word Count	Average Time Taken										
Birmingham	5,098	1,929	12.9%	15.6%	28.5%	0.1%	71.3%	4.2%	11.7%	7.1%	3.1%	2.8%	1.5	2.4	191.4	52.6										
Boole	3,455	1,593	22.3%	21.1%	43.5%	0.1%	58.9%	9.0%	16.1%	8.9%	3.6%	3.4%	2.4	3.5	202.3	50.5										
Bristol	2,591	825	14.0%	23.8%	37.7%	0.0%	59.9%	11.3%	17.3%	5.1%	3.0%	3.5%	2.2	4.8	164.2	48.8										
Cardiff	2,544	1,036	12.7%	21.8%	34.6%	0.2%	64.0%	5.5%	16.0%	9.3%	2.9%	2.3%	2.4	4.6	200.8	48.9										
Scotland	4,433	1,670	14.1%	19.3%	33.4%	0.1%	66.0%	6.7%	13.7%	6.1%	3.8%	3.7%	1.8	3.2	198.9	48.4										
Leeds	4,604	1,656	16.8%	22.0%	38.9%	0.0%	61.9%	8.1%	18.3%	6.7%	3.3%	2.7%	2.0	3.4	192.2	45.9										
Manchester	4,172	1,505	13.4%	24.7%	38.1%	0.1%	62.2%	8.4%	17.4%	6.4%	2.8%	2.7%	2.2	4.2	196.2	50.7										
Newcastle	3,035	1,072	13.6%	14.2%	27.8%	0.1%	73.0%	4.2%	10.2%	5.4%	3.8%	3.3%	1.4	2.3	162.3	49.2										
Nottingham	5,069	2,222	13.9%	22.7%	36.7%	0.0%	63.3%	9.9%	17.5%	4.4%	2.4%	2.5%	2.5	4.1	232.6	48.1										
Croydon	2,734	894	16.9%	24.3%	41.3%	0.1%	58.1%	11.1%	20.7%	4.8%	3.1%	2.3%	2.2	4.4	219.5	45.6										
Wembley	7,427	2,840	12.2%	19.1%	31.2%	0.1%	67.4%	6.8%	16.7%	5.6%	3.1%	1.3%	2.2	3.4	193.6	49.3										
National	45,162	17,342	14.5%	20.5%	35.0%	0.1%	64.8%	7.5%	15.9%	6.1%	3.0%	2.6%	2.1	3.6	196.5	49.5										

Column four is the percentage of people placed in the support group. The support group is for the most severely sick or disabled people, those whom taxpayers would unquestioningly deem ‘deserving’. What column four shows is that there is an expected ‘norm’ for the proportion of people who can be placed into the support group.

An Atos Healthcare spokesperson said:

“there are absolutely no targets set by either the DWP or Atos Healthcare for decisions on those found fit for work. Atos Healthcare does not make decisions on an individual’s entitlement to benefit and will likely never know what the outcome is.

“We do audit all practitioners on a random and rolling basis to ensure quality and consistency. One way we are able to trigger an audit process is to look at practitioners whose work is significantly outside of the average expected ranges over a period of time. Where this happens we look at the quality of their work and if no issues are found, no action is taken.

“This is certainly not about changing outcomes of assessments nor is it about reprimanding our practitioners in any way. It is about us checking our quality and making sure you can expect a consistent approach to your assessment whoever carries it out.”

However, whistle-blowers (both identified and anonymous) have provided testimony to contradict these claims; they explain that the audit practice is used to attempt to influence assessors to alter the results of their assessment. Dr Greg Wood is on record explaining it was exactly that demand from an auditor (that he alter his clinical opinion in a report) which led him to leave Atos and speak out in public. There are also accounts of the audit system being used in a punitive and bullying manner by middle managers in several different regions. Whilst the official policy at Atos is clear from their statement, it is also clear that, in practice, problems with audit and its use are widespread.

Each assessor is expected to see eight claimants each day and, as is abundantly clear from the table, the assessor can only put approximately 14.5% of those claimants, just one person of the eight of those assessed face-to-face each day, into the support group without triggering what assessors describe as a punitive audit process.

That 14.5% is an extrapolated forecast from estimated national trends. Using national trends to audit at a national level is perfectly acceptable practice – but this part of the audit system uses those same estimated national trends to also audit the performance of regions and crucially of individual assessors. Put more simply, it takes no account of variation – in an average day an assessor might see eight people with minor lower limb injuries expected to heal within six months, but they might equally see eight people with cancer, or awaiting a lung transplant or with severe learning disabilities - people who are so severely sick or disabled there is no way they are fit for work or work-related activity. The norms also ignore demographic variations in health and levels of disability – so for example, regardless of whether the assessor is in leafy Surrey where more people have good health or in inner city Glasgow, both those assessors are held to an identical ‘statistical norm’. This means significant disadvantage both to those claiming in areas where levels of sickness and disability are considerably higher than others and to the assessors working in those areas.

So what does happen if the assessor has eight people in a day who all need to be in the support group? These are ‘statistical norms’, not technically targets, but fear of audit, described by those working in Atos as ‘being put on the naughty step’, drives the behaviour of the humans being audited. It’s a complicated process which in practice means the norms act as a de facto target

system. Dr Greg Wood, who used to work as an assessor for Atos, blew the whistle on some of these bad practices earlier this year – when he described being asked by an auditor to change a report he had written to comply with the process of audit. Dr Wood commendably refused to behave in such an unprofessional manner and spoke out. The evidence now available validates Dr Wood’s claims.

In the month for which data is available (from earlier this year), column four clearly demonstrates that only one MSC (regional ‘medical examination centre’) has gone above the accepted norms as it is highlighted in red. That centre is Bootle and they are breaching the norm imposed by DWP because in that particular month the assessors in the centres in the Bootle region placed 22.3% of all those people they assessed into the support group. Out of all those people called for a face-to-face WCA examination, the assessment of 22.3% as so disabled or sick that they need to receive unconditional support does not seem like a huge proportion. However, from this table it is abundantly clear that it is too high a proportion for the audit system to allow. And that norm or percentage to be adhered to is imposed onto Atos by the DWP.

MSC	Below Normal Range		20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
	Total Number of Assessments	Total Number of Referrals	% In Support Group	% Curtailed	% Support Group Or Curtailed	% NFD	% 3m Prognosis	% 6m Prognosis	% 12m Prognosis	% 18m Prognosis	% 2y Prognosis	% Longer Term Prognosis	Average Physical score	Average Mental Score	Average PSS Word Count	Average Time Taken			
Birmingham	5,098	1,929	12.9%	15.6%	28.5%	0.1%	71.3%	4.0%	11.7%	7.1%	3.1%	2.6%	1.5	2.4	191.4	52.6			
Bootle	3,455	1,593	22.3%	21.1%	43.5%	0.1%	58.9%	9.0%	16.1%	8.9%	3.6%	3.4%	2.4	3.5	202.3	50.5			
Bristol	2,591	825	14.0%	23.8%	37.7%	0.0%	59.9%	11.3%	17.3%	5.1%	3.0%	3.5%	2.2	4.8	164.2	48.8			
Cardiff	2,544	1,036	12.7%	21.8%	34.6%	0.2%	64.0%	2.5%	16.0%	8.3%	2.9%	2.3%	2.4	4.6	200.8	48.9			
Scotland	4,433	1,670	14.1%	19.3%	33.4%	0.1%	66.0%	6.7%	13.7%	6.1%	3.8%	3.7%	1.8	3.2	198.9	46.4			
Leeds	4,604	1,655	16.8%	22.0%	38.8%	0.2%	61.9%	8.1%	18.3%	5.7%	3.3%	2.7%	2.0	3.4	192.2	45.9			
Manchester	4,172	1,505	13.4%	24.7%	38.1%	0.1%	62.2%	8.4%	17.4%	6.4%	2.8%	2.7%	2.2	4.2	196.2	50.7			
Newcastle	3,035	1,072	13.6%	14.2%	27.8%	0.1%	73.0%	4.2%	10.2%	5.4%	3.8%	3.3%	1.4	2.3	152.3	49.2			
Nottingham	5,069	2,222	13.9%	22.7%	36.7%	0.0%	63.3%	8.9%	17.5%	4.4%	2.4%	2.5%	2.5	4.1	232.6	48.1			
Croydon	2,734	894	16.9%	24.3%	41.3%	0.1%	58.1%	11.1%	20.7%	4.8%	3.1%	2.3%	2.2	4.4	219.5	55.6			
Wembley	7,427	2,840	12.2%	19.1%	31.2%	0.1%	67.4%	6.8%	16.7%	5.6%	3.1%	1.3%	2.2	3.4	193.6	49.3			
National	45,162	17,342	14.5%	20.5%	35.0%	0.1%	64.8%	7.5%	15.9%	6.1%	3.0%	2.6%	2.1	3.6	196.5	49.5			

Column five refers to the percentage ‘curtailed’ – here ‘curtailed’ means those placed into the Work Related Activity Group (WRAG), the group of people who may theoretically be capable of some form of work in the future with the right support and employment conditions. This group includes many people with progressive conditions such as Parkinson’s or Multiple Sclerosis. It is very clear from the highlighted cells in this column that the percentage of claimants the DWP will allow to be placed into the WRAG without triggering an audit is very low, and that there is no national consistency in the application of these standards. For example, Manchester have managed to trigger scrutiny - by placing 24.7% of those they have assessed within the month into the ‘curtailed’ or WRAG group, which is far above the national average of 20.5% - whereas both Birmingham and Newcastle have breached the norms by only placing 15.6% and 14.2% respectively into the WRAG, which is below the national average of 20.5%. Whilst many assessors have reported problems with punitive audit, both on and off the record, there has never been a report of punitive auditing practices triggered by being too far below the norm. This is how assessors who are genuinely performing poorly (as opposed to constantly being audited for making judgements that breach the norms) are supposed to be identified.

Column six is of particular interest; this is the audited norm for the total percentage of people allocated to the support group and ‘curtailed’ combined, i.e. all those who have reached the threshold applied by the DWP to the descriptors entitling them to ESA. In this particular month the national average for all those being granted the benefit is 35%, uncannily close to the one third of people Freud originally anticipated, in his 2007 report, should be deemed sufficiently sick or disabled to be eligible for support.

Column seven displays the ‘non-functional’ descriptors (NFD) – these are the descriptors the assessor is told to use when the person being assessed doesn’t fit within the WCA framework of descriptors normally used to assess. They allow for a claimant to be deemed to be not fit for work if undertaking work would pose risks for their health. ‘NFD’ applies to people with conditions as diverse as high risk behavioural issues, an inability swallow due to illness or (in women) high risk pregnancies. The ‘NFD’ requires a professional assessment of risk by a doctor or nurse. There has been considerable effort by campaigners to improve the usage of these descriptors, but it is all too clear from this column why those efforts have had such little impact; the national average allowed for the use of these descriptors is just 0.1%, just one person in every thousand assessed. Where these non-functional descriptors are used by assessors they have to complete the full ESA examination before they are able to apply this descriptor to the claimant.

MSC	Below Normal Range		50%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
	Total Number of Assessments	Total Number of Referrals	% In Support Group	% Curtailed	% Support Group Or Curtailed	% NFD	% 3m Prognosis	% 6m Prognosis	% 12m Prognosis	% 18m Prognosis	% 2y Prognosis	% Longer Term Prognosis	Average Physical Score	Average Mental Score	Average PSS Word Count	Average Time Taken			
Birmingham	5,098	1,929	12.9%	15.6%	28.5%	0.1%	71.3%	4.0%	11.7%	7.1%	3.1%	2.8%	1.5	3.4	191.4	52.6			
Bosley	3,455	1,593	22.3%	21.1%	43.5%	0.1%	58.9%	9.0%	16.1%	8.9%	3.6%	0.4%	2.4	3.5	202.3	50.5			
Bristol	2,591	825	14.0%	23.8%	37.7%	0.0%	59.9%	11.3%	17.3%	5.1%	3.0%	0.5%	2.2	4.8	164.2	48.8			
Cardiff	2,544	1,036	12.7%	21.8%	34.6%	0.2%	64.0%	6.5%	16.0%	9.3%	2.9%	2.3%	2.4	4.6	200.8	48.9			
Scotland	4,433	1,670	14.1%	19.3%	33.4%	0.1%	66.0%	6.7%	13.7%	6.1%	3.8%	0.7%	1.8	3.2	198.9	46.4			
Leeds	4,604	1,666	16.8%	22.0%	38.8%	0.8%	71.9%	8.1%	18.3%	6.7%	3.3%	2.7%	2.0	3.4	192.2	45.9			
Manchester	4,172	1,505	13.4%	24.7%	38.1%	0.1%	62.2%	9.4%	17.4%	6.4%	2.8%	2.7%	2.2	4.2	196.2	50.7			
Newcastle	3,035	1,072	13.6%	14.2%	27.8%	0.1%	73.0%	4.2%	10.2%	5.4%	3.8%	0.3%	1.4	2.3	152.3	49.2			
Nottingham	5,069	2,222	13.9%	22.7%	36.7%	0.0%	63.3%	9.9%	17.5%	4.4%	2.4%	2.5%	2.5	4.1	232.6	48.1			
Croydon	2,734	894	16.9%	24.3%	41.3%	0.1%	58.1%	11.1%	20.7%	4.8%	1.1%	2.3%	2.2	4.4	219.5	55.6			
Wembley	7,427	2,840	12.2%	19.1%	31.2%	0.1%	67.4%	6.8%	16.7%	5.6%	2.1%	1.9%	2.2	3.4	193.6	49.3			
National	45,162	17,342	14.5%	20.5%	35.0%	0.1%	64.8%	7.5%	15.9%	6.1%	3.0%	2.6%	2.1	3.6	196.5	49.5			

Columns eight to thirteen demonstrate why there is such a problem with ‘revolving door’ assessments. This has been one of the biggest causes of stress and misery to those very sick and disabled people recalled for reassessment every six to twelve months. The audit system demands that the majority of people being assessed receive a prognosis of three months when they score ‘sub-threshold’. The phrase ‘sub-threshold’ is automatically generated by the system whenever a claimant is awarded descriptors which score fewer than 15 points. In practice this score means they are found ‘fit for work’ by DWP decision makers who make the final decision about fitness for work based upon information gathered by the Atos assessors and any other evidence available in relation to the claimant. So, the ‘three month prognosis’ actually refers to people who have not scored enough points to be eligible for ESA and are assessed as ‘fit for work’. The system only permits 7.5% of those assessed to be given a prognosis (or re-assessment date) of six months, 15.9% is the average for a twelve month prognosis, 6.1% for eighteen months, 3.0% for two years and 2.6% for longer than two years (in effect, three years). Again we see this rough breakdown of one-third of people allocated benefit and two-thirds refused benefit reflected throughout the ‘statistical norms’. The sums of money wasted in reassessing people to a timetable imposed by an estimated audited norm, regardless of their actual health condition, should horrify every taxpayer.

Moving on to columns fourteen and fifteen, we see why people are receiving such low point scores at assessment. The assessors’ decisions are audited to an incredibly low point score, which effectively drives human behaviour and is the most likely reason for auditors to request the kind of changes reported by Dr Greg Wood – that he reduce the allocated descriptor point score of six he had awarded to a claimant (as medically appropriate) to one of zero (as required by the auditor). This auditing process does not apply to the tribunal services, however, leaving them free to make sensible, legally sound decisions based upon their professional interpretation of the legislation as guided by precedent. It requires 15 points overall scored from the descriptors (or criteria) for a

claimant to be placed in the Work Related Activity Group (WRAG) and 15 points on any one descriptor for the Support Group (SG). The small number of points a frontline assessor is able to give without eventually triggering audit clearly demonstrates why there is such a vast disparity between initial assessment and appeal; an audit may be triggered if the assessor awards the points outlined in the descriptors, i.e. if they go above an average of 2.1 points for physical conditions and 3.6 points for mental health conditions.

MSC	Below Normal Range		20%		20%		20%		20%		20%		20%		20%		20%		20%	
	Total Number of Assessments	Total Number of Referrals	% In Support Group	% Curtailed	% Support Group Or Curtailed	% NFD	% 3m Prognosis	% 6m Prognosis	% 12m Prognosis	% 18m Prognosis	% 2y Prognosis	% Longer Term Prognosis	Average Physical score	Average Mental Score	Average PSS Word Count	Average Time Taken				
Birmingham	5,098	1,929	12.9%	15.6%	28.5%	0.1%	71.3%	4.6%	11.7%	7.1%	3.1%	2.8%	1.6	2.4	191.4	52.6				
Bristol	3,455	1,553	22.3%	21.1%	43.5%	0.1%	59.9%	9.0%	16.1%	8.9%	3.6%	3.4%	2.4	3.5	202.3	50.5				
Cardiff	2,544	1,036	12.7%	21.8%	34.6%	0.2%	64.0%	6.5%	16.0%	9.3%	2.9%	2.3%	2.4	4.6	200.8	48.9				
Leeds	4,433	1,670	14.1%	19.3%	33.4%	0.1%	66.0%	6.7%	13.7%	6.1%	3.8%	3.7%	1.8	3.2	198.9	46.4				
Manchester	4,604	1,656	16.8%	22.0%	38.8%	0.0%	61.9%	8.1%	16.3%	6.7%	3.3%	2.7%	2.0	3.4	192.2	45.9				
Newcastle	4,172	1,505	13.4%	24.9%	38.1%	0.1%	62.2%	8.4%	17.4%	6.4%	2.8%	2.7%	2.2	4.2	196.2	50.7				
Nottingham	5,069	2,222	13.9%	22.7%	36.7%	0.0%	63.3%	9.9%	17.5%	4.4%	2.5%	2.5%	2.5	4.1	232.6	48.1				
Wembley	2,734	894	16.9%	24.3%	41.3%	0.1%	58.1%	11.1%	20.7%	4.8%	3.1%	2.3%	2.2	4.4	219.5	55.6				
National	7,427	2,840	12.2%	19.1%	31.2%	0.1%	67.4%	6.8%	16.7%	6.6%	3.1%	3.3%	2.2	3.4	193.6	49.3				
National	45,162	17,342	14.6%	20.5%	35.0%	0.1%	64.8%	7.5%	15.9%	6.1%	3.0%	2.6%	2.1	3.6	196.5	49.5				

Column sixteen, the 'average PSS word count', refers to the personal statements introduced after the first Harrington review, which were supposed to make it clearer to both the claimant and DWP decision maker why they have or have not been given the descriptors, and therefore points, required for receipt of benefit¹⁴. This was intended to bring some sense of the individual into the process - to humanise the explanation - but the focus for audit is whether the number of words falls within a particular range, rather than the accuracy of the statement, how consistent the statement is, or if it is easily understood by the Decision Maker at the DWP or the claimant. In practice assessors report that the purpose of the word count supervision was to keep the PSS short and therefore have less impact on productivity.

The final column, column seventeen, is the average time taken to perform an assessment. Human beings, sickness and disability are complicated both to explain and to be understood by an assessor. A genuine assessment of fitness to work should take as long as necessary, not be audited to fit within an approximate 50-minute window. Hence the eight-a-day minimum is onerous and detrimental to both the claimant and the practitioner.

Conclusion

As indicated throughout this article, the evidence gleaned from the original contract between Atos and DWP, testimony from Atos employees and Freedom of Information requests clearly indicates that outcomes for individual sick and disabled ESA claimants are not solely driven by the severity of their condition or the nature of their disability. On the contrary, as Lord Boswell predicted in 2007, the imposition of 'statistical norms' onto the Work Capability Assessment gears the outcome of the whole system to achieve the desired result - in effect, a cap on the overall number of people the system will permit to be eligible for Employment and Support Allowance.

The use of these 'statistical norms' as a management tool within Atos Healthcare can also be a secondary driver of the outcome for each individual person; although official policy is clear - that audit is not supposed to be used as a punitive process - it is also clear from whistle-blower evidence

¹⁴ <https://www.gov.uk/government/publications/work-capability-assessment-independent-review-year-1>

that, as in most large organisations, practice often does not equate to policy. Given clear reports that audit has indeed been used in a punitive manner by managers, there is a serious risk that the outcome for the claimant – whether they are placed in the Support Group and given unconditional support because they are unable to work, in the Work Related Activity Group on the basis that they are likely to be able to return to work in the future, or denied ESA altogether and expected to actively seek employment – may actually be driven more by the severity of **other** claimants' conditions than their own, particularly those claimants assessed on the same day by the same assessor!

It is also clear and, in the light of this evidence, totally unsurprising that this iniquitous system is causing immense distress, hardship and increased ill-health for the very people who most need support, whilst at the same time costing the taxpayer many millions of pounds in unnecessary assessments and appeals.

This evidence and analysis¹⁵ must be understood by politicians of all parties, since they alone can fix this cruel, wasteful and immoral system and prevent more unnecessary suffering. If sick and disabled people are to receive the support they need - and which a civilised, compassionate Western society is expected to provide - this article is essential reading for both Parliamentarians and the public alike.

¹⁵ **Note:** the analysis and conclusions explained in this article have been derived from evidence from a variety of sources, including individuals whose situation compels them to speak out anonymously. Whilst there is no absolute guarantee that the analysis is correct in every detail, advice has been sought from a number of statisticians and other relevant experts. On the evidence available, it is difficult to reach conclusions other than those set out below.