JOINTLY APPLYING A WORK SIMULATOR AND ATOM TO PREVENT OCCUPATIONAL ACCIDENTS AND MSD THROUGH WORKFORCE SELECTION

Lajos Izsó
Budapest University of Technology and Economics
izso.lajos@gtk.bme.hu

Blanka Berényi
Károli Gáspár University of the Reformed Church in Hungary
berenyi.blanka.pszi@gmail.com

Máté Pusker
CIVIL Plc
mate.pusker@gmail.com

Summary

Background and Aims: The goal of our work is the presentation of a particular – scientifically well-established – concept aiming to predict the propensity of individual job candidates for causing or suffering workplace accidents, and also for MSD-type *(Musculoskeletal Disorder)* occupational diseases, by further processing the performance parameters obtained by a work simulator (like ErgoScope) with the help of ATOM.
Methods: After introducing the problems of workplace accidents and MSDs, and critically reviewing the basic literature related to the so-called “work sample tests” and work simulators, the application possibilities of a specific, general-purpose work simulator, the ErgoScope, are presented for our purposes. After that, the possibilities of adequately integrating the ErgoScope and ATOM are described with particular respect to workplace accidents and MSDs, illustrated through a fictitious but realistically specified example.
Conclusions: The purposeful combination of the ErgoScope work simulator with ATOM can have a “synergistic” effect that reinforces each other’s effects, contributing to a significant reduction in the likelihood of workplace accidents and MSDs. Simply put, we propose to apply the appropriate outputs of the ErgoScope work simulator as inputs to ATOM.
Keywords: workplace accidents, occupational illnesses, MSD, work sample test, work simulator, workforce selection, ErgoScope, ATOM

DOI: 10.17627/ALKPSZICH.2023.3.75

The objectives of this article

The central message of this article is that – if the necessary methodological care is provided – the further processing of performance parameters obtained by a work simulator, with the help of ATOM, might result in a much better prediction of the propensity of individual job candidates for causing or suffering workplace accidents, and also for developing MSD-type *(Musculoskeletal Disorder)* occupational diseases.

The objective is to present this concept shortly, but still as informatively as possible. The tools to achieve this are the following.

- Providing an introduction to the problems of workplace accidents and MSD in Europe, since this is the application field of the proposed approach.
- Giving a concise review of the psychometric properties of work sample tests, since these form the basis for work simulators as workforce selection tools.
- Showing the main functionalities of work simulators (including the ErgoScope) tangentially.
- Having made these preparatory steps above, working out a fictitious, but realistic OSH (Occupational Safety and Health) example for the combined use of the ErgoScope and ATOM.
- Finally, using this fictitious example as a model, discuss the further possibilities and limits of this concept.

The following sections correspond to these points, while in the concluding discussion, an attempt is made to build a scientifically well-established construct for the concept of applying appropriate outputs of the ErgoScope work simulator as inputs to ATOM. Here we will argue why it is worth applying this concept in practice, and some of our related short-term plans are also outlined.

Introduction to the problems of workplace accidents and MSD

Regarding the recent statistics on workplace accidents in Europe (EU-28), EUROSTAT Statistics Explained (2022) provides the following rather gloomy data. In 2020

a) there were 2.7 million non-fatal accidents that resulted in at least four calendar days of absence from work;
b) there were 3,355 fatal accidents (about 20% of them within the construction sector);
c) 44.1% of all non-fatal accidents, and 63.1% of all fatal accidents happened in construction, transportation and storage, manufacturing, agriculture, forestry and fishing sectors;
d) about 66.5% of the total non-fatal accidents involved men;
e) the two types of particularly common injuries were wounds and superficial injuries (26.8% of the total).

Regarding the state of affairs in the field of work-related MSDs in Europe (EU-28), the European Agency for Safety and Health at Work (2023) provides the following shocking pieces of information. MSDs are the most prevalent occupational disease at the European level. Data from self-reporting through surveys *(European Working Condition Survey, European Health Interview Survey, European Labour Force Survey, European Survey of Enterprises on Emerging Risks)* inform us about the following: (1) three out of every five workers complained of MSDs; (2) more often than not, MSDs are accompanied by other health problems; (3) more than a third of workers reported that their work affects their health negatively; (4) 60% of workers with work-related health problem mentioned MSDs as most serious; (5) MSD prevalence is higher among older workers, (6) MSD prevalence decreases with education level.

With the rapid spread of modern information and communication technologies, mental work has become the main field of work for psychological and ergonomic research, while research on physical work has temporarily been neglected. Although the proportion of occupations requiring classic heavy physical work (e.g., miner, loader, earth-worker, material handler, etc.) has decreased radically, this still leaves a smaller number of such jobs. On the other hand – somewhat unexpectedly, and specifically in connection with IT-related jobs – it turned out that even office occupations, traditionally thought of as easy work, can often be physically demanding.

Not only high physical exertions may cause health risks, but also certain physical arm or hand operations – that in themselves, performed only once or several times can be considered as "easy" –, however, if repeated for a large number of times (up to tens of thousands during a work shift!). Examples include prolonged use of a computer keyboard, repeated execution of assembly sub-operations, frequent hand bending, gripping, twisting, squeezing, etc.

Because of the above, in many workplaces, instead of or in addition to the usual performance criteria related to success in the job (work performance), it is reasonable to raise "accident-free" and/or "MSD-free" work to the level of the performance criterion utilizing some reasonable quantification.

Work sample tests

Matching the most important characteristics of a person and a job is essential for job satisfaction and work efficiency. To ensure this, work and organizational psychology have developed several workforce selection approaches. One of them is the simulation of various work situations according to some critical, selected aspects, during which the behaviour of the candidate applying for the given job, is observed and evaluated in a standardized way. This is the work sample test. The advantages and disadvantages of applying such work sample tests are excellently summarized by HR-GuideSurvey.com (2023, opening screen of the link), therefore below we quote its most important parts:

- Main advantages: high reliability and content validity, difficulty for applicants to fake, and use of the same or similar equipment that is used on the job.
- Main disadvantages: costly to administer; and have less ability to predict performance on jobs where tasks take longer time (days or weeks).

Schmidt and Hunter (1998), and later Roth, Bobko and McFarland (2005) carried out large-scale meta-analyses on work sample test validity, and they found that compared with other of the studied procedures for predicting job success, the highest reported validity was for work sample tests (work simulators were not studied directly). The studied procedures were, among others and in increasing order of corresponding biases: work sample tests, integrity tests, conscientiousness tests, employment interview (structured), employment interview (unstructured), job knowledge, peer ratings, reference checks, job experience (years), biographical data measures, ACs (assessment centres), years of education, interests, age, etc.

These findings support the idea, that the use of work sample tests, including work simulators, is a good solution – despite its relatively high cost – in all areas where the consequences of wrong selection decisions could be quite serious. As Izsó (2012) put it, the jobs in which the risk of occupational accidents and MSDs is high, definitely can be considered such an area.

It has to be mentioned, that the ACs also operate partly on the work sample (work simulation) principle, but usually without using simulators, as special hardware and/or software equipment. An AC is a process, where candidates are given carefully designed specific tasks – mainly in the form of work samples – and are evaluated on their ability to perform a particular job. The ACs are strictly job-specific, for example, the book of Hale (2010) is about ACs specifically for selecting police and fire personnel.

Special target devices – called work simulators, to be defined in the next section – can create such simulated work situations of higher fidelity. In addition to the initial application of work simulators for purely aptitude testing (that is, for diagnostic purposes), the use of these devices recently also appeared for developing and improving job-relevant skills (that is, for training purposes).

Work simulators

While simulation, in general, is the imitation of a situation, environment, procedure or process, the simulator is a target device suitable for implementing the simulation itself.

Ergonomics, which according to its brief definition, is a human-centred technological design, deals with the optimization of different Man-Machine-Environment (MME) systems (Hercegfi & Izsó, 2007).

By definition, *ergonomic pathological factors* result from the structure of the MME system, the specific nature of the flow (exchange) of material, energy and information between man and machine, and also between man and the environment, as a result of physical, mental or emotional stress on the person as pathological effects of stress. MSDs are largely caused by *ergonomic pathological factors*, additional basic knowledge about this topic can be found in the publications of Béleczki et al. (2010) and Izsó (2011).

One particular type of simulation is in which a real, "flesh and blood" human gets into interaction with the Machine and/or Environment subsystems of a particular MME system. In what follows, we only deal with such simulations and the related simulators that by definition, are called "work simulators".

A good work simulator behaves largely similarly to the corresponding real Machine objects in terms of essential characteristics when interacting with humans. The degree of this similarity is characterized by *fidelity (realism).*

The *fidelity* of a work simulator, according to its general definition, is the measure of the accuracy of the simulation implemented with the given simulator. It measures how closely the given device follows the evolution (i.e., the behaviour) of the simulated situation, environment or process over time. One of the first reviews of terms, definitions and concepts related to the fidelity of practical work simulators was carried out by Hays (1980), who found that the use of the term (or wording) was not entirely uniform. He found a relative agreement that there are three main types or aspects of simulation fidelity:

- fidelity in external (physical) appearance (at the highest level, "photorealistic");
- functional fidelity (based on a model and related to operation/behaviour);

- psychological fidelity (refers to the sense of reality).

The work simulators operate on the work sample principle and are used mainly for workforce selection and training purposes. Although selecting candidates for given jobs based on their work sample performance goes back many centuries, or even a millennium, the first well-documented and systematic application of this principle is attributed to Hugo Münsterberg. In 1912, he successfully used a railway simulation method for selecting trolley operators first in Boston, and later on in other cities in the USA. Since then, the selection method of simulated work tasks (work sample) quickly spread.

Work simulators appeared in aviation as early as the 1930s. The Link Trainer was one of the first flight simulators, a very simple mock-up plane, designed to train pilots to operate basic flight controls. This later was followed by more and more sophisticated flight simulators, and nowadays already the big majority of civil and military plane types have their own high-fidelity, full-scale training simulators.

Another pioneer area in applying work simulators was the nuclear power industry. By the 1970s fully functional control room simulators had been developed for the most important reactor plant types of that time. The interested reader can find many details on this topic in the book of Skjerve and Bye (2011). The first author of this article has also been involved in developing simulator training methods for the nuclear power industry: Antalovits and Izsó (1999; 2003), Izsó (2001).

In the last several decades, many other vehicles, heavy machine, construction/mining equipment etc. simulators have been developed (e.g., CKAS [2023], TECH-LABS [2023], Caterpillar Inc. [2023], THOROUGHTEC Simulation [2023], CMLABS [2023]), not to speak of sophisticated simulators for military training purposes.

The best way to focus on our main interest presently, the general-purpose work simulators capable of assessing physical abilities, is to refer to the meta-analyses of Gouttebarge et al. (2004). These authors conducted their systematic literature search targeting the four most widely used work simulators (Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen Work System) in five databases (CINAHL, Medline, Embase, OSH-ROM and Picarta) using the keywords "functional capacity evaluation", "reliability" and "validity". They found that although the interrater reliability and predictive validity of the Isernhagen Work System were evaluated as good, the evaluation procedure used was not rigorous enough to allow any valid conclusion. Concerning the other three tools, neither convincing validity nor reliable data were found. These authors concluded that more rigorous studies are needed to demonstrate the reliability and validity of these work simulators.

Since another important work simulator, the Baltimore Therapeutic Equipment (BTE) had been left out from the review by Gouttebarge et al. (2004), a short evaluation of it will be presented here separately. The first important publication on the reliability and validity of BTE came out already more than three decades ago. The authors of this article – Kennedy and Bhambhani (1991) – determined the test-retest reliability and criterion validity of the BTE in an experiment involving 30 male volunteers. These volunteers acted as warehouse goods loaders and performed real (criterion) and simulated handwork. The three criterion tasks were done at light (CL), medium (CM), and heavy (CH) levels of intensity and three corresponding simulated tasks also were done at these three levels of intensity (SL, SM, SH). All of these tasks were repeated in a subsequent session. The authors experienced significant test-retest reliability concerning the two selected physiological parameters (oxygen consumption and heart rate). Although criterion-simulation correlation coefficients were also significant, consistently high criterion validity was found only at CL-SL (for oxygen consumption r = .81 and .83; for heart rate, r = .88 and .95).

Later some additional important details were published on the reliability and validity of BTE, e.g., Bhambhani, Esmail and Brintnell (1994), Ting et al. (2001).

The Ergoscope work simulator

The general-purpose ErgoScope work simulator, a new Hungarian development, is fitting to the progressive line of the Blankenship system, the Ergos work simulator, the ErgoKit, the Isernhagen Work System, and the Baltimore Therapeutic Equipment, and is free from some limitations of these antecedents. The ErgoScope shares the highest similarity with the Ergos and the Baltimore Therapeutic Equipment. The reason why the ErgoScope has been developed was mainly practical: the possibilities of taking into account special domestic needs, availability of quick and flexible service when needed, and detailed documentation in Hungarian.

As with all work simulators, the ErgoScope also simulates the "Machine subsystem" of the MME system corresponding to various work processes and activities. During ErgoScope simulator sessions, essential conclusions can be drawn about the observed person's physical, perceptual-thinking, and – by observation, to a limited extent – emotional characteristics too. *Figure 1* shows that the ErgoScope equipment consists of three standalone workstations (so-called panels) with different functions, which can be operated independently.

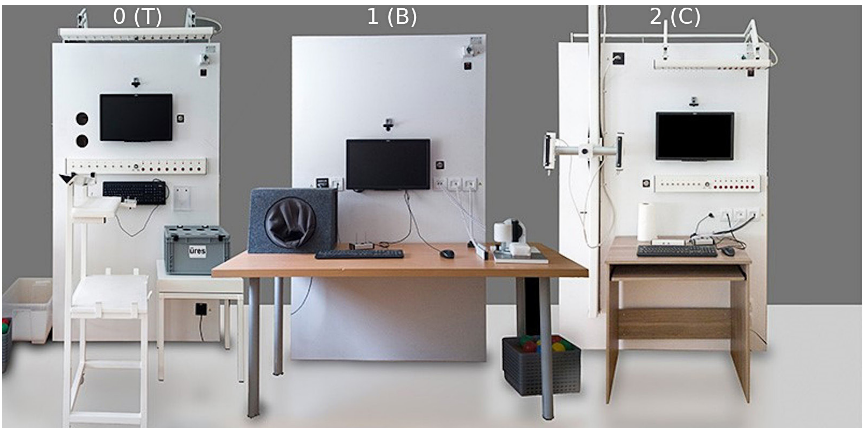

Figure 1. The three panels of the ErgoScope work simulator
Source: https://www.innomed.hu/munkaszimulatorok/

Tasks are performed using various measuring devices connected to data-collecting units, which transmit the measured data to a built-in computer that processes these data.

Panel 0 (T): *Static and dynamic force measurements* (using a bracket, movable on a vertical path):

- Static force measurements (static push/pull horizontally/vertically with two hands)
- Dynamic strength measurements (dynamic lifting to chair/shelf height with two hands, tools for dynamic measurement: scales, chest with weights)

Panel 1 (B): *Examining work performed while sitting*:

- Measurement of grip strength (fist grip with right/left hand, key grip with fingers of right/left hand, 3-point grip with finger of right/left hand, wrist flexion/extension/pronation/supination with right/left hand)
- Touch (with right/left hand)
- Keyboard operation (with right/left hand, with two hands with right/left sign)
- Pencil use ("pencil" use with right/left hand)

Panel 2 (C): *Examining capacity and monotony tolerance* (supplemented by examination of turning, switching and button pressing at chest height and overhead):

- Work capacity (moving crates, sorting balls, rolling balls)
- Monotony tolerance (tray moving, ball sorting, tray scrambling)
- Rotation (rotating knobs with the dominant hand from the eyes/overhead)
- Use of switches (use of switches face-to-face/overhead)
- Use of push-buttons (use of push-buttons directly/overhead)

Altogether 215 concrete-specific objective performance parameters can be measured on these three panels in 36 measurement modes (elementary simulated work situations). The following two main types of work diagnostic surveys are distinguished:

- Full job diagnostic survey: for career guidance, this survey is usually carried out when the client has no concrete ideas about his future job or has several competing ideas but cannot choose. We can often offer the client jobs in which they could perform exceptionally or at least well in terms of the skills required for that job. In the cases of weaker performances, we can recommend targeted skill development. If development is not possible, the search for other, better-fitting jobs follows.
- Targeted job diagnostic survey: this survey is usually carried out when the client comes with a specific job idea or the future employer requests objective data about the client's applicability.

Izsó, Székely and Dános (2015) studied the specific possibilities of this work simulator, especially for use as aptitude testing of people with altered workability, and also touched on its skills development possibilities. (Izsó [2015] compiled a methodological manual, in which the recommended reference values for the ErgoScope parameters can be found.)

Various forms of occupational accidents, MSDs resulting from incorrect/inaccurate limb and full-body movements, and inappropriate exertion can be prevented by properly using the ErgoScope as a professional aptitude testing tool. The parameters measured by the ErgoScope can only be used as predictors of successful future work – i.e., being free from workplace accidents and MSDs – if:

1. these can be considered relevant for the given job based on the knowledge and experience of OSH specialists;
2. a suitable database is available for these parameters for reference purposes.

Regarding the first question, the answer is based on the expertise of OSH specialists. Regarding the second one, we currently have an ErgoScope database, built from the measured parameters of 297 healthy and 100 disabled people. Since installing the first pieces of ErgoScope in 2016, we have participated in many related projects and gained considerable experience in application methodology. The use of ErgoScope parameters as input data (predictors) for the ATOM software package is very promising. As the central message of this article, this problem area is outlined separately in the next section.

Applying data obtained by the Ergoscope as inputs to ATOM for reducing the risk of workplace accidents and MSDs

Requirements for properly combining the ErgoScope and ATOM

As described in more detail in the preceding articles of this special issue – Izsó; Gergely and Takács; Pusker, Gergely and Takács – ATOM, developed by us, is an AI-based expert system for predicting job success based on suitable *predictors* and relevant *success criteria* of the given the job.

A predictor in this context is a variable suitable to predict the future job success of applicants.

Predictors can typically include, among many others: qualifications, relevant work experience, job-specific skills (e.g., driving license, computer proficiency, ability to speak a particular language), certain test scores, objective parameters measured by electro-mechanic or computerized aptitude-testing devices or work simulators, etc.

> Important comment: Since MSDs and occupational accidents often stem from false/imprecise limb and whole-body movements or inappropriate strength exertions, OSH professionals must take into consideration this viewpoint while selecting ErgoScope performance parameters as predictors for a given job.

The *job success criteria* – again, among other things – can typically be:

- actual quantitative and/or qualitative production data (however, such data – for theoretical or practical reasons – are not available for many jobs);
- management's scores on the employee's performance (the disadvantage of these is that they are generally not statistically reliable enough, primarily due to the so-called "halo effect" and the "leniency" and "severity" biases).
- Important comment: in the following fictitious OSH example the long-term accident or/and MSD-freeness must be properly operationalized (quantified) to be used as job success criteria.

If such criteria are available and appropriate – strongly correlated – predictors can also be found for them, based on these predictors, the person's success in the given job can be predicted with a high probability.

ATOM can be applied if valid predictors are available for at least about 100 employees who have already proven to be successful in a given job to different extents (including also failure). This rough practical rule of thumb of using minimally about 100 data points, is based on our experiences gained during targeted ATOM studies.

Providing the required data ATOM's competing and flexible learning algorithms "learn" the relationships between the predictors (as input variables) and the job success criteria (as output variables). Based on the resulting model, ATOM can predict the (expected) success of new applicants for the given job under investigation based only on the values of the predictors (Gergely & Takács, this special issue).

Regarding jobs where the risk of workplace accidents or/and MSD is high, the professionally correctly designed combination of the two domestically developed leading technologies (the ErgoScope with its broad range of functions and the resulting potentially high content validity, and ATOM with its extreme flexibility) is expected to have a strong synergistic effect. This process can be formulated generally with the following simple "IF X, THEN Y" type rule.

1. IF, for a specific job, the OSH specialists determine the parameters (predictors) that can be measured by the ErgoScope and are considered relevant concerning the risk of workplace accidents, or MSD;
2. and these predictors are measured with the help of the ErgoScope in the cases of at least about 100 employees already working in the given position, whose job success in terms of being free from workplace accidents and MSDs are known and numerically different;
3. and after that, the experts use the ATOM's machine learning (ML) algorithms to build the model that best describes the relationship between predictors and job success based on this data;
4. THEN, examining the candidates newly applying for the given job by the ErgoScope, the future job success of these applicants, defined as being accident- and/or MSD-free, can be predicted with relatively high accuracy.

A fictitious OSH example for combining the ErgoScope and ATOM

A goods loader (counter loader) fills shelves and loading areas and keeps the goods clean and tidy in grocery stores, shops, and other wholesale units. For this kind of job, the following ErgoScope performance parameters can be considered relevant:

- Panel 0 (T): static pull / push horizontally / vertically / dynamic lift to chair height/ shelf height;
- Panel 2 (C): work endurance (complex task sequence: moving crates, sorting balls, rolling balls), monotony tolerance (complex task sequence: moving trays, sorting balls, rolling balls).

Let us assume that the ErgoScope performance parameters given above (as predictors) are available for 150 successful and 50 unsuccessful workers for this job, as well as the degree of their actual job success on a five-point scale (its value is 2, 3, 4 or 5 for the successful and 1 for the unsuccessful). By loading this data into ATOM, the learning algorithms "learn" the relationships between the predictors and the job success value given on this five-point scale (in this case, characterizing the persons' middle- or long-term accident or/and MSD-freeness). Finally, if the company wants to hire new employees for this job, then the same ErgoScope performance parameters must be measured for these applicants. Having done so, using the model ATOM provided – based on the data of above mentioned 150 successful and 50 unsuccessful workers – ATOM can predict the probabilities of new applicants falling into each job success category (these are the so-called “labelling probabilities”). Job success (in our particular case, the persons’ accident or/and MSD-freeness), however, is also determined by certain psychological characteristics in addition to the physical (motor and force exertion-related) skills/abilities identified by the OSH professional and measured by the ErgoScope. Therefore, the results of appropriate psychological tests must also be used as predictors, but we will not deal with this problem here (only briefly in the *Discussion*).

In short, in our example, by entering the predictors (the corresponding ErgoScope performance parameters) for the new applicants into ATOM, we get the probabilities with which applicant may fall into each of the five categories of the applied job success scale.

The way of applying ATOM’s results depends on the current labour supply and demand situation. When there is labour oversupply, applicants must be sorted according to the decreasing “expected probability of job success” first within the 5 (best) success grades. However, when there is a labour shortage (i.e., when, in principle, all applicants should be hired), hiring those with a 1 (worst) expected job success among new applicants is not recommended. The experience is that these people mainly result an extra expenditure for the company as they eventually either quit on their own or have to be fired.

However, if for some reason, the company is still forced to hire from among the applicants rejected by ATOM, then the applicants must be sorted according to the increasing “probability of job success” first within those with 1 (worst) job success grade. After that, applicants have to be selected from among those, who are relatively lower on this list, and they have to be assessed by traditional HR methods (job interview, overall impression shown at the interview, performance at previous workplaces, living and housing conditions, family situation, the orderliness of finances, etc.). Based on these, the company may override the results obtained from ATOM, but it has, of course, certain risks.

A rule of thumb is that ML models’, including ATOM’s, classification performance is acceptable if their hit rate is at least about 20% better than the hit rate by chance alone. The background of this guideline, for reasons of space, is not presented here. The actual hit rates (both overall, relating to all categories simultaneously, and also corresponding only to particular categories) can be calculated from the ATOM’s output files using the appropriate functionalities of the *Setup (Beállítások)* primary window (Pusker, Gergely & Takács, this special issue, *The four primary windows*).

In our fictitious example, we used a five-point job success scale, to which $1/5 = 0.2 \rightarrow 20\%$ random hit probability would correspond. If instead, ATOM provides a global forecast with at least 40% accuracy concerning all categories, that is already a significant surplus. However, if our success scale had only two levels (e.g., 0 = *“likely to fail”* and 1 = *“likely to succeed”*), then $1/2 = 0.5 \rightarrow 50\%$ would be the random hit probability, and this should be increased to at least 70%.

Presenting the proposed approach

During the workforce selection process, the following two main biases are distinguished usually. The first comes from the applicants' side, who are generally willing to pretend to be better than they actually are. This tendency, as Henle et al. (2019) published, might result in faked personality inventories and intentional fraud causing misinterpretation of resumés by HR personnel. The second concerns the applied methods' side – e.g., König and Langer (2022) – since most selection methods involve human decisions, usually by HR personnel, that are inherently error-prone.

We, however, claim that there also exists a third main bias. The source of this relates to the basic question *"Have we chosen the best data processing procedure from among the many possible ones in terms of given input-output relationships?"*. This bias does not relate to data but stems from the chosen data analysing methods.

The first main bias remains henceforward also in the cases of AI-supported workforce selection methods, like ATOM, while the second one can be reduced by applying appropriate AI-driven methods. Reducing the third main bias, however, is only possible if a proper variety of procedures are used, either sequentially or simultaneously, and the results of the best-performing one are accepted. Although this approach requires increased computational resources, it is already quite feasible using today's quick computers. Notwithstanding, we have not found in the literature AI-based methods operating on this principle. Our ATOM system, however, is based on this novel principle: it runs simultaneously many ML algorithms and the outputs of the "winner" (the best performing one) are considered as results (for more details refer to Gergely and Takács: this special issue). The main advantage of competing algorithms is that they can adapt to the diversity of workplace selection, training data of varying size and quality, expert evaluation, and the specific characteristics of the job and latent data generation processes. Thus, our ATOM system can effectively reduce this third type of distortion too. Furthermore, if ATOM uses properly chosen outputs of the ErgoScope work simulator as predictors, this combination hopefully results in relatively bias-free predictions (refer to *Table 1*).

Table 1. A conceptual comparison of hypothesized resultant biases for different combinations of data gathering and data analysing procedures for job success prediction (provided that a simple additive summation rule is valid).

Data analysing procedures	**Data gathering procedures**					
	1. Work sample tests, including work simulators (e.g., ErgoScope)	2. Questionnaire-based methods (conscientiousness tests, integrity tests, etc.)	3. Interview-based methods (structured / unstructured)	4. Peer ratings	5. ACs	6. Biographical measures (years of education / employment, etc.)
1. ATOM	2	3	4	5	6	7
2. Other AI-based methods	3	4	5	6	7	8
3. Traditional statistical methods	4	5	6	7	8	9
4. Traditional non-statistical HR methods	5	6	7	8	9	10

Source: based on own research data

The serial numbers of the data gathering and data analysing methods are at the same time the ranks of the corresponding biases. Thus, concerning data gathering, “work sample tests, etc.” the first column has the smallest, while the last column “biographical measures, etc.” has the greatest biases. Similarly, concerning data analysis, “ATOM” has the smallest, while “traditional non-statistical HR methods” has the biggest biases. The numeric fields contain the sum of ranks concerning biases corresponding to the procedures in the respective columns and rows. The smaller these sum ranks are, the better (the more bias-free) the corresponding are combinations of the “data gathering” – “data analysis” methods.

The first of the above-mentioned three main biases, attributable to applicants, usually occurs at data gathering procedures 2. and 3. (The biases at procedures 1., 4., 5., and 6. are caused by other factors.)

The second bias, attributable to HR personnel, usually occurs during data gathering procedure 3. (The biases at procedures 1., 2., 4., 5., and 6. are caused by other factors.)

The third bias, attributable to the choice of data processing methods, might occur in all four data analysing procedures, but its magnitude is probably the minimum in the case of ATOM. This is a strong, scientifically well-established argument for using the outputs of the ErgoScope as predictors fed to ATOM.

The accuracy of predictions depends largely on the quality of the input data, as the popular adage says, “garbage in, garbage out”. If the algorithms are trained with low-quality data, then the classification result will also be of poor quality. Analyses with low-quality data can raise serious validity problems, but to a certain extent, these can be compensated by using different, more robust statistical procedures (Gergely & Vargha, 2021), as it is done in ATOM.

Since data quality is a multi-dimensional concept, in data science different authors have identified roughly 6–16 distinct dimensions for different purposes. Of these, the first 6 basic dimensions that most publications – e.g., Wang et al. (2002), Batini and Scannapieca (2006), Lee et al. (2006) – are in alignment with. The following two of those are especially relevant to us here: *accuracy* and *relevancy.*

Accuracy: is a measure of how well the data reflects the object being described along the given characteristics (How well does the data reflect reality, irrespective of the relevancy to the actual *matter studied*?). This dimension corresponds to the earlier mentioned first and second main biases.

Relevancy: is a measure of the level of consistency between the content of data and the studied areas of interest (in our case, the job success). In other words, it is the extent to which data answers the question of the actual study (To what extent are the data applicable and useful for predicting job success?). Data relevancy means different things for different task contexts: what is relevant for predicting success in a particular job, may not be relevant for other purposes.

Here we go back to the fictional case of selecting candidates for the goods loader job and consider a bit more closely some steps and circumstances of the combined application of the ErgoScope work simulator and ATOM. In the very first step work psychologists and OSH experts – based on their earlier experiences and overall expertise – compile a set of possible predictors consisting of certain personality traits; cognitive, perceptual, motor and force exertion functions. This can be taken as the first iteration step made by human expertise, to be followed by many other computational steps to be done by the concurrent algorithms of ATOM. These starting decisions on the predictors to be applied are decisive since even the best algorithms are later confined by them.

Suppose that the intensity levels of these chosen predictors, minimally necessary for acceptable job performance, are known empirically from the company's earlier workforce selection campaigns. *Table 2* shows this in simplified form: in the "Characteristics" column the chosen predictors are listed, while in the four "Level of Characteristics" columns the minimal requirements are indicated by bold solid polygonal chain lines in percentile units. In the same four columns the actual values of three hypothetical candidates can also be found similarly by dotted, dashed and dotdash chain lines. Suppose that all these data are *accurate* enough.

Table 2. Comparing the fictitious requirements for the goods loader job with the actual values of hypothetical candidate 1, 2 and 3 in terms of competence characteristics.

Group of characteristics	Measuring instruments	Characteristics*	LEVEL OF CHARACTERISTICS (in percentiles**)			
			0%	25%	50%	75% 100%
Personality traits	Suitable personality tests	Scale 1				
		Scale 2				
		Scale 3				
		Etc.				
Cognitive functions	Suitable cognitive tests	Cogn. function 1				
		Cogn. function 2				
		Cogn. function 3				
		Etc.				
Perceptual functions	Electronic measuring devices	Perc. function 1				
		Perc. function 2				
		Perc. function 3				
		Etc.				
Motor functions**	**ErgoScope work simulator**, special measuring devices	**Moving hutches**				
		Handgrip				
		Wrist stretching				
		Etc.				
Force exertion functions**	**ErgoScope work simulator,** special measuring devices	**Horizontal push**				
		Vertical pull				
		Dynamic lifting				
		Etc.				
Other groups of characteristics as necessary	To be determined…	To be identified…				

Note:

* These characteristics are considered relevant to different degrees for this job. These scales are used as predictors of future job success, and – for simplicity reasons – all are of positive polarity ("the bigger is the better" type).

** A percentile is the percent of cases that are at or below a score.

*** To these functions concrete characteristics (performance parameters) examples are indicated that can be measured by the ErgoScope work simulator.

Requirements by a given hypothetical job

Values of hypothetical candidate 1

Values of hypothetical candidate 2

Values of hypothetical candidate 3

Source: edited on the basis of own research data

The percentiles are proper units for both the minimal job requirement and the actual values of candidates since these correctly reflect the fact that if a predictor value is very infrequent in the population of possible candidates, that very predictor has very high predictive power. This job, as seen in *Table 2*, requires such a high handgrip value that about 75% of the population cannot produce. We can also see that hypothetical candidate 3 is able to exert handgrip that about 90% of the population cannot do. On the contrary, the requirement concerning scale 2 personality trait is only about a 25% percentile, which about 75% of the population can perform.

Concerning hypothetical candidate 1, we can see that while in the cases of personality traits, cognitive and perceptual functions, the values of characteristics are above the minimal requirements of the job, in the cases of motor and force exertion functions the values are below the minimum requirements. This data set is *relevant* since contains the appropriate motor and force exertion characteristics (it is another question that since these characteristics are below the required level, these decrease the success probability in the goods loader job).

Concerning hypothetical candidate 2, we can see that while in the cases of personality traits, cognitive and perceptual functions, the values of characteristics are above the minimal requirements of the job, in the cases of motor and force exertion functions the values are missing. This data set is *irrelevant* since this only contains such characteristics that have little or almost nothing to do with the success of the goods loader job. This fact represents a lack of information concerning the success probability in the goods loader job.

Concerning hypothetical candidate 3, we can see that in the cases of all characteristics, the values are above the minimal requirements of the job. This data set is *relevant* since contains the appropriate motor and force exertion characteristics (and since all these characteristics are above the required, these increase the success probability in the goods loader job).

Psychology, as a pure theoretical science, primarily wants to explain psychic phenomena with the simplest and most parsimonious models possible, while placing less emphasis on prediction. The consequence is that the results can only be generalized within a closed theoretical framework and often have negligible predictive power (Robinaugh et al., 2021). In contrast, ML algorithms (especially deep neural networks) aim to maximize the prediction accuracy of the models, and mostly they do not provide an understandable explanation for how the phenomenon actually works (Yarkoni & Westfal, 2017).

Therefore, when we started developing ATOM for applied work and organizational psychological purposes, at the same time we also decided in favour of maximizing the prediction accuracy, and based on this, maximizing the efficiency of practical workforce selection decisions. The price we have to pay for it is that we will not necessarily know which variables and to what extent played a role in the outcome.

These limitations have certain consequences concerning *Table 2*. This table, as it is, has mainly didactic goals and therefore its content and the related interpretation above are rather simplified.

Although work psychologists and OSH experts naturally can compile a valid set of possible predictors, in reality they can never determine in advance the intensity levels of these chosen predictors, in concrete numerical terms, that are minimally necessary for acceptable performance in a given job. The reason for it is that ATOM's ML algorithms, optimized for maximum prediction accuracy, hardly provide any information about how the predictors are actually interacting (increasing or decreasing each other's effects), and consequently, how much is the resultant predictive power of the individual predictors. So, it could still happen, that a predictor thought rightfully very relevant by a human expert, turns out to be seemingly unimportant because of the confusing complex interactions between the many predictors. This is especially true if the number of predictors is high (say several hundred).

Discussion

From the EUROSTAT Statistics Explained (2022) publication, we have learnt that although in Europe significant progress has already been made in the field of OSH in recent decades, still more than 3,300 fatal and about 2.7 million non-fatal accidents occur in the 28 EU member states every year. The most prevalent occupational diseases still are MSDs: three out of every five workers complained of MSDs in the last years. These facts justify why preventing workplace accidents and MSD-type occupational diseases is regarded as a primary goal nowadays.

As mentioned earlier in the *Work sample tests* section, the meta-analyses on work sample tests revealed that compared with other procedures for predicting job success, the highest reported validity was for work sample tests. Therefore, an effective way for preventing workplace accidents and MSDs could be to develop workforce selection methods targeting specifically these problems based on appropriate work simulators, which are purposeful implementations of carefully selected work sample tests. We can sum up that a professionally appropriate combination of the use of the ErgoScope work simulator and the capabilities of ATOM may result in a "synergistic" effect, reinforcing each other's effects thus contributing to a further reduction in the occurrence of workplace accidents and MSDs. Therefore, applying appropriate outputs of the ErgoScope work simulator as inputs to ATOM is proposed.

In a recent OSH conference – Izsó (2022) – we announced our plan to realize the fictitious example of the goods loader job, discussed above, in the near future in the form of a large-scale field study. Similarly, in the longer term, we also plan to carry out job success prediction studies by the combined use of the ErgoScope and ATOM involving other physically demanding jobs.

Összefoglalás

Munkaszimulátor és az ATOM együttes alkalmazása munkabalesetek és MSD típusú foglalkozási megbetegedések megelőzésére a munkaerő kiválasztása útján

Háttér és célkitűzések: Munkánk célja annak a – tudományosan jól megalapozott – koncepciónak a bemutatása, amely szerint egy adott fizikai munkakörre jelentkező munkavállaló hajlama munkabaleset előidézésére vagy elszenvedésére, illetve MSD *(Musculoskeletal Disorder)* típusú foglalkozási megbetegedésre jól előrejelezhető, ha prediktorként az ATOM rendszerben munkaszimulátorral (pl. ErgoScope-pal) nyert releváns mérési adatokat használunk.
Módszer: A munkabalesetek és MSD típusú foglalkozási megbetegedések problémakörének általános ismertetése, valamint az ún. „munkaminta tesztekkel" és munkaszimulátorokkal kapcsolatos szakirodalom kritikai áttekintése után egy konkrét általános célú munkaszimulátor, az ErgoScope alkalmazási lehetőségeit vizsgáljuk jelenlegi céljaink kapcsán. Ezt követően annak a lehetőségeit vizsgáljuk meg – egy fiktív, de realisztikus példával illusztrálva –, hogy hogyan lehet a legelőnyösebb módon integrálni az ErgoScope és az ATOM rendszereket a munkabalesetek és az MSD típusú foglalkozási megbetegedések lehető legpontosabb előrejelzése érdekében.
Következtetések: Az ErgoScope és az ATOM együttes alkalmazása egyfajta „szinergikus" hatást eredményezhet, amely felerősíti a két rendszer külön-külön történő alkalmazásának a hatásait, és ez jelentősen hozzájárulhat a munkabalesetek és az MSD típusú foglalkozási megbetegedések valószínűségének csökkenéséhez. Egyszerűen fogalmazva, azt javasoljuk, hogy az ErgoScope-pal nyerhető, szakszerűen kiválasztott mérési adatokat (az ErgoScope alkalmas kimeneteit) alkalmazzuk az ATOM rendszer bemeneteiként.
Kulcsszavak: munkabaleset, foglalkozási megbetegedés, MSD, munkaminta teszt, munkaszimulátor, munkaerő-kiválasztás, ErgoScope, ATOM

References of this Special Issue

Izsó, L. (2023). The concept of an AI-based expert system (ATOM) for predicting job success. *Alkalmazott Pszichológia*, *25*(3), 5–13.
Gergely, B., & Takács, Sz. (2023). ATOM – a flexible multi-method machine learning framework for predicting job success. *Alkalmazott Pszichológia*, *25*(3), 15–30.
Pusker, M., Gergely, B., & Takács, Sz. (2023). ATOM's structure – employee and employer feedback, survey site. *Alkalmazott Pszichológia*, *25*(3), 53–72.

References

Antalovits, M., & Izsó, L. (1999). Self-assessment and learning in nuclear power plant simulation training. In J. Misumi, R. Miller, & B. Wilpert (Eds.), *Nuclear Safety: A Human Factors Perspective* (pp. 243–256). Taylor & Francis.

Antalovits, M., & Izsó, L. (2003). Assessment of Crew Performance and Measurement of Mental Effort in a Cognitively Demanding Task Environment. In Hockey, G. R. J., Gaillard, A. W. K., & Burov, O. (Eds.), *Operator Functional State* (pp. 284–290). NATO Science Series, IOS Press.

Batini, C., & Scannapieca, M. (2006). *Data Quality. Concepts, Methodologies and Techniques.* Springer-Verlag.

Béleczki L., Varga J., Plette R., & Ungváry Gy. (2010). Ergonómiai alapismeretek. Ergonómiai kóroki tényezők okozta megbetegedések. In Ungváry Gy., & Morvai V. (Eds.), *Munkaegészségtan. Foglalkozás-orvostan, foglalkozási megbetegedések, munkahigiéné* (pp. 711–718; 3. átdolgozott és bővített kiadás). Medicina Könyvkiadó Zrt.

Bhambhani, Y., Esmail, S., & Brintnell, S. (1994). The Baltimore Therapeutic Equipment work simulator: Biomechanical and physiological norms for three attachments in healthy men. *American Journal of Occupational Therapy, 48*(1), 19–25. 10.5014/ajot.48.1.19. PMID: 8116780, https://www.semanticscholar.org/paper/The-Baltimore-Therapeutic-Equipment-work-simulator%3A-Bhambhani-Esmail/5c787f8937148587c244e4f6ee8d21b81950e67d

Caterpillar Inc. (2023, January 21). *CAT simulators*. https://catsimulators.com/videos

CKAS (2023, January 21). *Driving Simulators*. https://www.ckas.com.au/driving_simulators_29.html?gclid=EAIaIQobChMI4oXbpMf2_AIV0EeRBR1QegLvEAAYASAAEgKH6_D_BwE

CMLABS (2023, January 25). *Construction Equipment Training Simulators*. https://www.cm-labs.com/immersive-simulation-products/construction-equipment-training-simulators/

European Agency for Safety and Health at Work (2023, February 11). *Work-related diseases*. https://osha.europa.eu/en/themes/work-related-diseases

European Agency for Safety and Health at Work (2023, February 10). *Work-related musculoskeletal disorders: prevalence, costs and demographics in the EU*. Publications Office. https://data.europa.eu/doi/10.2802/66947

EUROSTAT Statistics Explained (2022, December 11). *Accident at work statistics*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics

Gergely, B., & Vargha, A. (2021): How to Use Model-Based Cluster Analysis Efficiently in Person-Oriented Research. *Journal for Person-Oriented Research*, *7*(1), 22–35.

Gouttebarge, V., Wind, H., Kuijer, P. P. F. M., & Frings-Dresen, M. H. W. (2004). Reliability and validity of Functional Capacity Evaluation methods: A systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen Work System *International Archives of Occupational and Environmental Health, 77*(8), 527–537 https://doi.org/10.1007/s00420-004-0549-7

Hale, C. D. (2010). *The Assessment Center Handbook for Police and Fire Personnel*. Charles C. Thomas Publisher Ltd.

Hays, R. T. (1980). *Simulation Fidelity: A Concept Paper*. U. S. Army, Research Institute for the Behavioral and Social *Sciences*.

Henle, C. A., Dineen, B. R., & Dulffy, M. K. (2019): Assessing intentional resume deception: Development and nomological network of a resume fraud measure. *Journal of Business and Psychology, 34*(1), 87–106.

Hercegfi K., & Izsó L. (Eds.) (2007). *Ergonómia*. BME tankönyv. Typotex Kiadó.

HR-GuideSurvey.com (2023). Personnel Selection: Methods: Work Sample Tests. https://hr-guide.com/Selection/Work_Sample_Tests.htm

Izsó, L. (2001). Fundamentals of the Model Behind the COSMOS Methodology Used for Team Assessment in Simulator Training. *Journal of Occupational Safety and Ergonomics*, *7*(2), 163–178.

Izsó L. (2011): Ergonómiai tényezők vizsgálata a munkahelyen. In Ungváry Gy. (Ed.), *Munkahigiénés gyakorlatok*. OMFI.

Izsó L. (2015): *Az ErgoScope munkaszimulátor által mérhető paraméterek javasolt indulási referencia-értékei*. Methodological Handbook. InnoMed.

Izsó L. (2022): *A munkaképesség felmérése munkaszimulátor alkalmazásával, a beválás előrejelzése mesterséges intelligencia segítségével.* Magyar Üzemegészségügyi Tudományos Társaság XI. Kongresszusa. 2022. október 6. http://www.mutt.hu/anyagok/MUTT_kongresszus_2022_Esztergom_Programfuzet.pdf

Izsó L., (2012): Munkaszimulátorok alkalmazásának lehetőségei a munkavégzés biztonságának javításában. *Munkavédelem* és *Biztonságtechnika*, *2012*(4), 10–16.

Izsó, L., Székely, I., & Dános, L. (2015): Possibilities of the ErgoScope high fidelity work simulator in skill assessment, skill development and vocational aptitude tests of physically disabled persons (*„Best Paper Award”* winner conference paper). 13th International Conference of the Association for the Advancement of Assistive Technology in Europe, Sept. 9–12, 2015. Budapest, Hungary. As book chapter: In Sik-Lányi, C., Hoogerwerf, E. J., Miesenberger, K., & Cudd, P. (Eds.). *Assistive Technology* (pp. 825–831). IOS Press.

Kennedy, L. E., & Bhambhani, Y. N. (1991). The Baltimore Therapeutic Equipment Work Simulator: Reliability and validity at three work intensities. *Archives of Physical Medicine and Rehabilitation*, *72*(7), 511–516.

König, C. J., & Langer, M. (2022): Machine learning in personnel selection. In Strohmeier, S. (Ed.), *Handbook of Research on Artificial Intelligence in Human Resource Management* (pp. 149–167). Edward Elgar.

Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2006), *Journey to Data Quality.* MIT Press.

Robinaugh, D. J., Haslbeck, J. M., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021): Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*(4), 725–743.

Roth, P. L., Bobko, P., & McFarland, L. A. (2005): A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58*(4), 1009–1037. https://doi.org/10.1111/j.1744-6570.2005.00714.x

Schmidt, F. L., & Hunter, J. E. (1998): The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2). 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Skjerve, A. B., & Bye, A. (2011): *Simulator-based Human Factors Studies Across 25 Years*. Springer.

TECH-LABS (2023, January 30). *Forklift Personal Simulator*. https://tech-labs.com/products/forklift-personal-simulator

THOROUGHTEC Simulation (2023, January 30). *Full mission simulators*. https://www.thoroughtec.com/cybermine-full-mission-mining-simulators/?gclid=EAIaIQobChMIprLdqMr2_AIVCNGyCh28zA9VEAAYASAAEgLWQfD_BwE

Ting, W., Wessel, J., Brintnell, S., Maikala, R., & Bhambhani, Y. (2001). Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men. *American Journal of Occupational Therapy*, *55*(2), 184–190. https://doi.org/10.5014/ajot.55.2.184

Wang, R. Y., Ziad, M., & Lee, Y. W. (2002). *Data Quality.* Kluwer Academic Publishers.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.