Title: The comparison of normative reference data from different gait analysis services

Abstract: Comparison of normative data between gait analysis services offers the potential to harmonise data collection protocols. This paper presents a method for such a comparison based on an assumption that the root mean square difference from the inter-service mean is a reflection of systematic differences in protocols and that the average standard deviation includes a component attributable to within-centre measurement variability.

Substantial normative datasets from two highly respected clinical services were compared. The RMS difference for the difference from the inter-centre mean was less than 1.7° for all kinematic variables apart from hip rotation (2.9°) and foot progression (2.1°), less than 0.1Nm/kg for all joint moments and than 0.21W/kg for all joint powers. The two centres showed very similar normative standard deviations.

The data demonstrates a high degree of consistency between data from two highly regarded gait analysis services and establishes a baseline against which other services can assess their performance. An electronic appendix includes data to facilitate this comparison.
THE COMPARISON OF NORMATIVE REFERENCE DATA FROM DIFFERENT GAIT ANALYSIS SERVICES

Ornella Pinzone¹, Michael H. Schwartz²,³, Pam Thomason⁴, Richard Baker¹

¹University of Salford, UK
²Gillette Children’s Specialty Healthcare, St. Paul, MN, USA,
³University of Minnesota, Minneapolis, MN, USA,
⁴Royal Children’s Hospital, Melbourne, Australia.

e-mail: O.Pinzone@salford.ac.uk
ABSTRACT

Comparison of normative data between gait analysis services offers the potential to harmonise data collection protocols. This paper presents a method for such a comparison based on an assumption that the root mean square difference from the inter-service mean is a reflection of systematic differences in protocols and that the average standard deviation includes a component attributable to within-centre measurement variability.

Substantial normative datasets from two highly respected clinical services were compared. The RMS difference for the difference from the inter-centre mean was less than 1.7° for all kinematic variables apart from hip rotation (2.9°) and foot progression (2.1°), less than 0.1Nm/kg for all joint moments and than 0.21W/kg for all joint powers. The two centres showed very similar normative standard deviations.

The data demonstrates a high degree of consistency between data from two highly regarded gait analysis services and establishes a baseline against which other services can assess their performance. An electronic appendix includes data to facilitate this comparison.
1. INTRODUCTION
In the past it has been considered acceptable for clinical gait analysis services to vary in their
data capture protocols and reference datasets were collected to allow for these differences [1-3]. As the clinical gait analysis matures, there is a growing requirement for standardization
between services [4, 5]. This has been underlined by two articles [6, 7] emphasising the
differences between laboratories in 3d gait analysis data, raising concern within the
orthopaedic community [8, 9]. The rationale for collecting reference datasets in the future
should thus be to harmonise protocols through comparison between different services. This
study describes a mechanism for such a comparison and illustrates this by comparing data
from two internationally regarded gait analysis services.

2. MATERIALS and METHODS
The normative reference data (means and standard deviations) in routine use at two gait
analysis services, (Gillette Children’s Specialty Healthcare, GCSH, and the Royal Children’s
Hospital, Melbourne, RCH) were compared. The normative reference data were created using
data from 81 patients, with an age range between 4 and 17 years at one centre, and 31
patients, with age between 6 and 17 years at the other centre. All data had been collected at
self-selected walking speed, with a Vicon kinematic measuring system (Oxford, UK) and
AMTI force plates (Watertown, MA, USA). The length of the walkways were respectively 8
m and 15 m. Knee Alignment Devices (KAD) were used in static calibration. Trajectories had
been filtered with a Woltring spline filter [10] and then processed using Plug-in Gait [11]
software (Vicon, Oxford, UK). Data were sampled to 51 values during the gait cycle;
however there is no particular reason to believe this method would be sensitive to this value.

Means \((m_{ijt}, i \text{ refers to service, } j \text{ to gait variable and } t \text{ to % of gait cycle})\) and standard
deviations \((s_{ijt})\) of the clinically important kinematic and kinetics variables from the two gait
analysis services were plotted together to visualize the level of agreement (Error! Reference source not found. and Error! Reference source not found.). Assuming that the two cohorts walked similarly, then differences in the mean measurements reflect systematic differences in measurement technique between the two centres. Although comparing one mean to the other appears an obvious choice for comparisons, each service is interested not in how it compares with the other, but how it compares with the true mean for the population. Given that there is no reason to suspect that one set of measurements is “better” than the other, the grand mean between the services \( M_{jt} \) is actually the best estimate of the true mean. Given that differences in technique are likely to be characteristic of the service rather than the participants, a simple mean was preferred to a mean weighted by number of participants.

\[
M_{jt} = \frac{1}{N} \sum_{i=0}^{N} m_{ijt}
\]

Where \( N \) is the number of services (2 in this case).

Systematic differences were then quantified by considering the difference between the mean for each service and the grand mean \( \Delta_{ijt} (= m_{ijt} - M_{jt}) \). This approach has the advantage that the method can be extended to the comparison of any number of services.

Three parameters are assumed to be of interest: \( RMS\Delta \), \( mean\Delta \), and \( SD\Delta \).

\[
RMS\Delta_{ij} = \sqrt{\frac{1}{n} \sum_{t=0}^{n} \Delta_{ijt}^2}
\]
\[
\text{mean}\Delta_{ij} = \frac{1}{n} \sum_{t=0}^{n} \Delta_{ijt}
\]

\[
SD\Delta_{ij} = \sqrt{\frac{1}{n} \sum_{t=0}^{n} \left( \Delta_{ijt} - \text{mean}\Delta_{ij} \right)^2}
\]

Where \( n \) is the number of time points across the gait cycle. The absolute maximum value of \( \Delta_{ijt} \) was also calculated. If only two services are being compared \( \Delta_{1jt} = -\Delta_{2jt} \) by definition, and only one set of results need to be reported. (Note also that \( \text{RMS}\Delta^2 = \text{mean}\Delta^2 + SD\Delta^2 \)).

The offset percentage (OP, with \( = 100 \cdot \frac{\text{mean}\Delta^2}{\text{RMS}\Delta^2} \)) was also calculated for kinematic parameters and represents the proportion of the mean square that arises from the fixed offset. This is useful because high values of OP will tend to indicate the direct effect of differences in protocols for the placement of markers on segments adjacent to the joint and in the plane the angle is measured in [4]. Low values will tend to indicate secondary effects of more distant markers or in a different plane. This distinction may be useful in investigating and correcting the source of discrepancies.

The standard deviations (\( s_{ijt} \)) represent the variability of measurement at each centre. They tend to be relatively constant over the gait cycle and the average value is thus taken as the representative measure for each variable. This is a combination of physiological variability and measurement error and the smaller this value, the more consistently a specific protocol has been applied. These values from different services are meaningful in this context in their own right (as opposed to the means \( m_{ijt} \), which are only meaningful in this context when compared) and are simply reported separately for each centre.
3. RESULTS
There were no statistically significant differences between body mass, height and leg length at the two centres. There were differences in absolute walking speed (p<0.0001) with patients from centre 2 walking 10% faster on average than patients from centre 1.

A qualitative analysis of the kinematics graphs (Error! Reference source not found.) shows that agreement is generally good. There are small differences in pelvic rotation, hip flexion and knee flexion with slightly larger differences in internal hip rotation, ankle dorsiflexion and foot progression during swing. The statistical parameters (Table 1) reflect these with the largest differences ($RMS\Delta$) of between 2° and 3° in hip rotation and foot progression. Kinetic data also show generally good agreement, but detailed scrutiny suggests some differences in hip and knee extensor moment data and all the power graphs.

4. DISCUSSION and CONCLUSIONS
This paper describes a quantitative method for comparing normative datasets for different gait analysis services. This has been illustrated with the comparison of two datasets but the method can be extended to as many datasets as desired. Differences are generally small with only two kinematic parameters having an $RMS\Delta$ above 2° which McGinley et al. [5] suggested as acceptable (although four parameters showed maximum differences of greater than 2°).

From Table 1, it can be seen that the two variables showing the largest offsets, hip rotation and foot progression, also show a large offset percentage suggesting a simple difference in marker placement protocol. Hip rotation is known to be highly dependent on the placement of knee alignment devices [12, 13], foot progression will be dependent on the alignment of forefoot and heel markers during the static calibration [4]. The process by which marker data
is incorporated into the calculation of kinetic variables is considerably more complex than for
the kinematic data and none of the offset percentages is particularly high. It is probable that
the offset percentage will be less useful in explaining the source of discrepancies between
datasets for the kinetic data as for the kinematic data.

The main limitation of the method is that it assumes that characteristics of the two reference
cohorts are similar and that differences reflect measurement technique. The age, weight,
height and gender match of both samples is good and factors such as socio-economic and
ethnic background (mainly Caucasian) are likely to be similar. There is some evidence of
differences in temporal and spatial parameters between specific ethnic groups [14] and this
should be considered as an additional source of variation if very different groups are being
compared.

There are differences in speed between the two cohorts and the small differences in pelvic
rotation, hip and knee dorsiflexion appear to be consistent with the speed related changes
reported by Schwartz et al. [15]. These differences are small though (in comparison to the
overall standard deviations) and controlling for speed is probably not required. There are also
small differences in timing events between the two approaches. This is likely to have a small
effect on the overall results, leading to a small overestimate in the true variability.

Differences in the standard deviations between the services are generally small. Whilst this is
reassuring, the way in which the physiological variability (p) and measurement variability
(m) combine to produce the overall variability (s = \sqrt{p^2 + m^2}) means that this measure will
be relatively insensitive to changes in measurement variability that are equal to or less than
the physiological variability. Specific studies [e.g. 16] are required to give confidence that
repeatability is acceptable.
The means and standard deviations of the two centres and the characteristics of the two reference cohorts are lodged as an electronic appendix allowing any centre to compare its measurements against these benchmarks.
Table 1: Summary data describing how data from one centre differs from combined data (kinematic data in degrees, moments in N·m/kg and powers in W/kg).

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Difference in mean measurements</th>
<th>Variability (Mean SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSΔ</td>
<td>MeanΔ</td>
</tr>
<tr>
<td>Pelvic Tilt</td>
<td>0.13°</td>
<td>-0.08°</td>
</tr>
<tr>
<td>Hip Flexion</td>
<td>1.06°</td>
<td>0.28°</td>
</tr>
<tr>
<td>Knee Flexion</td>
<td>1.59°</td>
<td>0.00°</td>
</tr>
<tr>
<td>Ankle Dorsiflexion</td>
<td>1.30°</td>
<td>-0.63°</td>
</tr>
<tr>
<td>Pelvic Obliquity</td>
<td>0.40°</td>
<td>0.02°</td>
</tr>
<tr>
<td>Hip Abduction</td>
<td>0.57°</td>
<td>0.08°</td>
</tr>
<tr>
<td>Pelvic Rotation</td>
<td>0.72°</td>
<td>0.04°</td>
</tr>
<tr>
<td>Hip Rotation</td>
<td>2.87°</td>
<td>-2.74°</td>
</tr>
<tr>
<td>Foot Progression</td>
<td>2.13°</td>
<td>1.88°</td>
</tr>
<tr>
<td>Hip Extensor Moment</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>Knee Extensor Moment</td>
<td>0.05</td>
<td>-0.03</td>
</tr>
<tr>
<td>Plantarflexor Moment</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Hip Abductor Moment</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Knee Abductor Moment</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Ankle Rotation Moment</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Hip Power</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>Knee Power</td>
<td>0.17</td>
<td>0.04</td>
</tr>
<tr>
<td>Ankle Power</td>
<td>0.21</td>
<td>-0.04</td>
</tr>
</tbody>
</table>
7. Figure(s)
7. Figure(s)
Fig. 1: Mean and standard deviations between GCSH (grey) and RCH (black). Kinematics normative reference data.

Fig. 2: Mean and standard deviations between GCSH (grey) and RCH (black). Kinetics normative reference data.
HIGHLIGHTS

- The data forms a baseline against which other services can assess their performance.

- A method for comparison of normative data between gait analysis services is presented.

- Normative datasets from two highly respected clinical services showed close agreement.

- Both sets of kinematic and kinetic data are available for benchmarking purposes.